

Automatische Erfassung präziser Trajektorien in Personenströmen hoher Dichte

Maik Boltes

Forschungszentrum Jülich GmbH
Institute for Advanced Simulation (IAS)
Jülich Supercomputing Centre (JSC)

Automatische Erfassung präziser Trajektorien in Personenströmen hoher Dichte

Maik Boltes

Schriften des Forschungszentrums Jülich
IAS Series

Band 27

ISSN 1868-8489

ISBN 978-3-95806-025-8

Bibliografische Information der Deutschen Nationalbibliothek.
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte Bibliografische Daten
sind im Internet über <http://dnb.d-nb.de> abrufbar.

Herausgeber
und Vertrieb: Forschungszentrum Jülich GmbH
Zentralbibliothek, Verlag
52425 Jülich
Tel.: +49 2461 61-5368
Fax: +49 2461 61-6103
E-Mail: zb-publikation@fz-juelich.de
www.fz-juelich.de/zb

Umschlaggestaltung: Jülich Supercomputing Centre, Forschungszentrum Jülich GmbH

Druck: Grafische Medien, Forschungszentrum Jülich GmbH

Copyright: Forschungszentrum Jülich 2015

Schriften des Forschungszentrums Jülich
IAS Series Volume 27

D 38 (Diss., Köln, Univ., 2014)

ISSN 1868-8489
ISBN 978-3-95806-025-8

Persistent Identifier: [urn:nbn:de:0001-2015011609](https://nbn-resolving.org/urn:nbn:de:0001-2015011609)

Alle Rechte vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (Druck, Fotokopie oder in einem anderen Verfahren) ohne schriftliche Genehmigung des Verlages reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Kurzzusammenfassung

Simulationen können helfen, Verkehrsanlagen für Fußgänger komfortabel und sicher zu gestalten. Das Verständnis über die Fußgängerdynamik ist dabei wesentlich für die Entwicklung verlässlicher Modelle. Hierfür sind detaillierte und reproduzierbare Daten realer Bewegungen von Menschenmassen und Individuen nötig, um das Bewegungsverhalten zu analysieren, daraufhin Modellideen zu entwickeln, deren Umsetzung zu kalibrieren und am Ende das Ergebnis zu validieren.

Diese Arbeit beschreibt die Erhebung solcher Bewegungsdaten, die aus Überkopfaufnahmen von Kameras gewonnen werden. Dabei werden individuelle Laufwege jeder Person extrahiert. Mit Hilfe dieser Trajektorien werden die wichtigsten Messgrößen zur Beschreibung der Dynamik von Menschenströmen wie u. a. die Personendichte, die Geschwindigkeit und damit der Personenfluss bestimmt. Besonders bei Stau und Gedränge ist mit kritischen Situationen zu rechnen. Daher ist es für die zur Extraktion der Laufwege entwickelten Verfahren wichtig, dass diese auch für hohe Personendichten verlässlich anwendbar sind.

Die Bewegung von Fußgängern wird von vielen Faktoren wie z. B. der Geometrie, der Personendichte, der Motivation und dem Kulturkreis beeinflusst. Zur Untersuchung dieser Vielzahl an Einflussfaktoren sind eine große Anzahl an Versuchen mit zum Teil hohen Personenzahlen durchzuführen. Die automatische Erfassung der Laufwege bietet im Vergleich zur manuellen Erfassung deutliche Vorteile hinsichtlich des Zeitaufwandes, der Genauigkeit und der Reproduzierbarkeit.

Die Prozesskette der Extraktion besteht aus der Kalibrierung der Bilder, der Detektion der Personen, dem Tracking und der Bestimmung des realen Aufenthaltsortes. Für die Detektion einer Person wurden verschiedene den unterschiedlichen Anwendungsfällen und örtlichen Gegebenheiten angepasste Marker mit entsprechenden Erkennungsverfahren entwickelt. Insbesondere für Feldstudien ist zudem ein markerloses Verfahren erarbeitet worden, das mit Hilfe von Stereokameras auch in dichten Menschenmassen eine hohe Erkennungsrate aufweist. Alle Entwicklungen sind in die den gesamten Prozess abbildende Software *PeTrack* eingeflossen. Für die Nachvollziehbarkeit der Ergebnisse wird neben den Verfahren auch die technische Umsetzung der Erfassung beschrieben.

Abstract

Simulations can help make facilities for pedestrians safer and more comfortable. A proper understanding of crowd dynamics is essential to developing reliable models for such simulations. Detailed and reproducible datasets of real crowd movements are needed for analysis and modelling. Such datasets are also required for later calibration and validation of said models.

This thesis describes the collection of such data from overhead video recordings. Individual trajectories are extracted and make it possible to obtain the most relevant quantities of the dynamic e.g. pedestrian density, velocity and flow. Traffic jams and other high density situations are of special interest since this is where critical situations are to be expected. Therefore the developed methods have to also reliably extract an individual's movement in such situations.

The movement of pedestrians is affected by many factors such as geometry, crowd density, motivation and culture. To investigate these numerous influences a large number of experiments with a huge number of participants have been carried out. The automatic extraction of the trajectories provides a significant advantage compared to manual methods in terms of the time required, accuracy and reproducibility.

The extraction process consists of the image calibration followed by the detection, tracking and determination of the real world position of all individuals. For the detection of a person various markers and corresponding extraction techniques have been developed for the different applications and local conditions. A markerless method was also developed, which is especially useful for field studies. Through the use of stereo cameras high detection rates were achieved without markers, even in high density situations. All developments regarding the extraction process have been integrated into the software *PTrack*. To enable a deeper understanding of the results the technical aspects of the trajectory collection are described in addition to the recognition techniques.

Danksagung

Die Experimente und Feldstudien sowie die Entwicklung der in dieser Arbeit vorgestellten Verfahren durfte ich als wissenschaftlicher Mitarbeiter am Jülich Supercomputing Centre des Forschungszentrums Jülich durchführen. An dieser Stelle möchte ich all jenen danken, die durch ihre fachliche und persönliche Unterstützung zum Gelingen der Arbeit beigetragen haben.

Zuerst möchte ich Prof. Dr.-Ing. Ulrich Lang danken, der mit großem Interesse die Arbeit begleitet und die Begutachtung übernommen hat. Prof. Dr. Andreas Schadschneider hat durch seine Unterstützung in allen Phasen der Arbeit maßgeblich zu ihrem Gelingen beigetragen. Er stand mir jederzeit mit ehrlichem Rat und hilfreicher Tat zur Seite. Prof. Dr. André Bresges danke ich für die Übernahme des Vorsitzes der Prüfungskommission.

Ein besonderer Dank gebührt Prof. Dr. Armin Seyfried, der mich fortwährend in meiner Arbeit bestärkt und dazu motiviert hat, diese in eine Dissertation einfließen zu lassen. Er hat mir sehr viel Vertrauen entgegengebracht und ich habe in seiner Abteilung ein außerordentlich kollegiales und inspirierendes Arbeitsumfeld vorgefunden. Besonders hervorzuheben ist die tolle Zusammenarbeit mit Stefan Holl, Dr. Bernhard Steffen, Dr. Jun Zhang, Dr. Mohcine Chraïbi, Dr. Ulrich Kemloh, Daniel Salden, Svenja Schmidt und Wolfgang Mehner, die mir sehr bei der Durchführung der Experimente und Analyse der Laufwege geholfen haben.

Meine Vorgesetzten Prof. Dr. Dr. Thomas Lippert, Dr. Norbert Attig, Dr. Rüdiger Esser und Dr. Herwig Zilken haben mich immer in meiner Arbeit unterstützt und Freiräume bei der Institutsarbeit für meinen Ausflug in den Bereich der Fußgängerdynamik gelassen.

Meiner Frau Elvira ist zu danken, dass Sie mir in wichtigen Zeiten den Rücken frei gehalten und in den übrigen Zeiten klar gemacht hat, dass die Familie ebenso wichtig wie der Beruf ist. Trotz alledem mussten meine Kinder insbesondere in den letzten zwei Jahren viel auf ihren Vater verzichten – ich gelobe Besserung! Zu guter Letzt sollen meine Eltern und Freunde nicht unerwähnt bleiben, die durch ihren Ansporn und das Verständnis für die vielen Absagen zum Gelingen dieser Arbeit beigetragen haben.

für Milena, Melissa und Marvin

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Ziel und Aufbau der Arbeit	4
2	Existierende Verfahren zur Lokalisierung von Personen	7
2.1	Abgrenzung zu ähnlichen Gebieten	8
2.2	Verfahren zur Erfassung der Position einer Person	11
2.3	Erfassungstechnik	15
2.3.1	Kamera	15
2.3.1.1	Hintergrundsubtraktion	18
2.3.1.2	Klecks-Analyse	18
2.3.1.3	Vorlagen-Abgleich	19
2.3.1.4	3D-Form-Modell	21
2.3.1.5	Detektion mit Vorwissen	22
2.3.2	Stereokamera	26
2.3.3	Mehrkamerasystem	31
2.3.4	Infrarotkamera	33
2.3.5	Wärmebildkamera	35
2.3.6	3D-Kamera	36
2.3.7	RGB-D-Sensor	37
2.3.8	Laufzeit-Kamera	39
2.3.9	Laserscanner	40
2.3.10	Weitere Systeme	41
2.3.11	Vergleich	43
2.4	Tracking	45

3	Erfassung von Laufwegen	47
3.1	Kalibrierung	48
3.1.1	Vermessung	48
3.1.2	Entzerrung	53
3.2	Detektion mit Marker	56
3.2.1	Marker mit Strukturelementen	61
3.2.2	Farbmarker	66
3.3	Tracking	69
3.3.1	Neuberechnung und Qualität	75
3.3.2	Extrapolation	76
3.3.3	Fehlbilder	77
3.3.4	Bereinigung	78
3.3.5	Verschmelzung	80
3.4	Positionsbestimmung	82
3.4.1	Farbkodierung	84
3.4.2	Disparität	87
3.4.3	Fehler der Disparität	96
3.5	Detektion ohne Marker	99
3.5.1	Hintergrundsubtraktion	102
3.5.2	Isolinien gleicher Disparität	104
3.5.3	Approximierende Ellipsen	109
3.5.4	Ellipsenpyramiden	109
3.5.5	Identifikation	111
3.6	Ergebnisse	119
3.6.1	Kalibrierung	121
3.6.2	Größenbestimmung mittels Disparität	123
3.6.3	Trajektorien mit Strukturelementen markierter Personen	127
3.6.4	Trajektorien mit Farbmarkern markierter Personen	132
3.6.4.1	Parameterwahl	133
3.6.4.2	Fehler durch perspektivische Sicht	135
3.6.4.3	Fehlerbetrachtung anhand virtueller Person	138
3.6.4.4	Glattheit der Trajektorien	140
3.6.5	Disparität zur Personenerkennung	141
3.6.6	Trajektorien aus Stereoaufnahmen	148
3.6.6.1	Fehler bei der Erkennung der Personen	148

3.6.6.2	Genauigkeit der Position	156
3.6.6.3	Glattheit der Trajektorien	162
3.6.7	Feldstudie	163
3.6.7.1	Vergleich der kommerziellen Systeme	166
3.6.7.2	Selbstentwickeltes Verfahren	168
3.6.8	Rechenzeit	169
3.6.9	Fehlereinfluss auf Maße der Fußgängerdynamik	174
3.6.9.1	Fehler bei der Erkennung	174
3.6.9.2	Ortsabhängiger Fehler	176
3.6.9.3	Ortsabhängige Fehlervarianz	177
3.6.9.4	Ungenauigkeit des Detektionsortes	177
3.6.10	Vergleich der Detektionsmethoden	178
4	Versuche	181
4.1	Durchführung	183
4.1.1	Erfassungstechnik	183
4.1.1.1	Verdeckung	187
4.1.1.2	Sichtfeld und überlappende Kameras	188
4.1.1.3	Brennweite und Sensorgröße	192
4.1.1.4	Belichtung und Beleuchtung	192
4.1.1.5	Bildwiederholrate	193
4.1.1.6	Sonstige Kameraeigenschaften	194
4.1.1.7	Übertragungsschnittstelle	195
4.1.1.8	Synchronisation	196
4.1.1.9	Speicherung	196
4.1.1.10	Montage	198
4.1.2	Probanden und Aufbauten	198
4.2	Auswertung	200
4.3	Messmethode	202
4.4	Anwendung	206
4.4.1	Engstelle	207
4.4.2	T-Kreuzung	211
4.4.3	Evakuierungsassistent	215
5	Zusammenfassung und Ausblick	217
5.1	Zusammenfassung	217
5.2	Ausblick	219

Literatur	224
A Teilveröffentlichungen	273
B Zeitschriften und Konferenzen	277
B.1 Zeitschriften und Bücher	277
B.2 Konferenzen und Schulen	278
C Videos	283
D Implementationsbeschreibung	289
D.1 Voraussetzungen	289
D.1.1 Hilfsmittel	289
D.1.2 Programmbibliotheken	289
D.2 Aufbau von PeTrack	290
D.3 Nutzung von PeTrack	292
D.4 Nutzung von combine	295
E Kommerzielle Produkte zur Personenerfassung	297
F Nomenklatur	301
F.1 Symbolverzeichnis	301
F.2 Einheiten	303
G Glossar	305

Abbildungsverzeichnis

1.1	Konzerte mit großen Menschenmengen	1
1.2	Heilige Moschee in Mekka während des Haddsch	2
1.3	Simulierte Evakuierungszeiten eines Korridors	2
1.4	Fundamentaldiagramm der Fußgängerdynamik	3
1.5	Weg vom Experiment zur Simulation	4
2.1	Optischer Fluss und Helligkeitsgradienten als Merkmale zur Personendetektion	7
2.2	Optischer Fluss zur Bewegungsanalyse von Menschenmassen	10
2.3	Bewegungserfassung des gesamten Körpers	10
2.4	Frühe Feldstudie zum Zusammenhang von Fluss und Geschwindigkeit von Greenshield	17
2.5	Veranschaulichungen von Verfahren zur Personendetektion	20
2.6	Fehlerraten aktueller Verfahren zur Personendetektion	22
2.7	Mit Kameras erfasste Experimente anderer Gruppen	24
2.8	Laufwegermittlung mittels mitgeführtem Tracking-System	26
2.9	Position in der Bewegungsebene	28
2.10	Tiefenbilder zur Personendetektion	29
2.11	Kamerabild projiziert in das Weltkoordinatensystem	29
2.12	Kugelstruktur in Punktwolke zur Kopflokalisierung	30
2.13	Verarbeitung des Tiefenfeldes zur anschließenden Leitkegelsuche	31
2.14	Hough-Transformation zur Positionsbestimmung bei Mehrkamerasystemen	32
2.15	Pixelvarianz zur Kopfdetektion im Mehrkamerasystem	33
2.16	Versuchsaufbau mit Infrarotkameras beim Pedigree-Projekt	34
2.17	Intensitätsbilder einer Wärmebildkamera	35
2.18	Bild einer überkopf angebrachten Wärmebildkamera	35
2.19	Tiefenfeld einer Kinect	38
2.20	Verfahren zur Personendetektion im Tiefenfeld einer Kinect	39
2.21	Laborplattform pamela	41

2.22	Laufweg durch ein Gebäude mittels einer inertialen Messeinheit	43
2.23	RFID-Technik zur Personenortung	44
3.1	Schema des Ablaufs der Erfassung der Laufwege	47
3.4	Koordinatentransformation zwischen Welt-, Kamera und Bildkoordinatensystem . .	48
3.2	Schritte zur Erfassung von Laufwegen	49
3.3	Schritte zur Erfassung von 3D-Laufwegen	50
3.5	Lagebestimmung eines Bildpunktes im Kamerakoordinatensystem	52
3.6	Wirkung der Verzerrungsparameter	54
3.7	Kalibrierungsmuster mit detektierten Merkmalspunkten	55
3.8	Liste bisher eingesetzter Marker	58
3.9	Liste bisher eingesetzter Marker (Fortsetzung)	59
3.10	Hintergrundsubtraktion in einer Bildfolge	61
3.11	Isolinien und approximierende Ellipsen	63
3.12	Isolinien und approximierende Ellipsen (Ausschnitt)	64
3.13	Schritte zur Erkennung von Marker Nr. 1	65
3.14	HSV-Farbraum und die Umsetzung der Farbwahl in PeTrack	66
3.15	Farbmarker zur Personenerkennung einschließlich Schließen und Öffnen der Maske .	67
3.16	Farbverteilung des Farbmarkers im HV- und HS-Diagramm	68
3.17	Tracking mit Subpixel-Genauigkeit	69
3.18	Vergleichsbereiche in Gauß-Pyramiden	70
3.19	Scheitern des Tracking bei starkem Schatten	71
3.20	Differenzbild von Folgebildern	73
3.21	Vergleich von Detektion und Tracking	74
3.22	Drift der Trajektorie durch Suche des dunkelsten Markerpixels	74
3.23	Wiederholtes Verfolgen mit höherer Qualität	76
3.24	Lücke in der Trajektorie bei fehlendem Bild in der Aufzeichnung	77
3.25	Geschwindigkeitsverteilung entlang der optischen Achse	79
3.26	Überlappende Trajektorien	80
3.27	Transformation von Trajektorien zur Angleichung von Kameraperspektiven	81
3.28	Ausschnitt von 3D Laufwegen zweier überlappender Kameras	81
3.29	Interpolation beim Verschmelzung von Trajektorien	82
3.30	Datenstrukturen der Laufwege in der Bildebene und im Weltkoordinatensystem . .	83
3.31	Ortsfehler aufgrund unterschiedlicher Personengrößen	84
3.32	Farben des Farbmarkerelementes von Marker Nr. 1	85

3.33 Farbdetektion entlang der Laufwege	86
3.34 Verteilung der Personengröße laut Befragung und automatischer Ermittlung	87
3.35 Einfluss der Basislänge auf Disparität	89
3.36 Rektifizierung versetzter Blickwinkel eines Stereokamerasystems	90
3.37 Rektifizierung	90
3.38 Disparitäten unterschiedlicher Verfahren	92
3.39 Farbkodierte Disparität	93
3.40 Geometrische Bestimmung der Disparität	94
3.41 Zusammenhang zwischen Punktentfernung zur Stereokamera und deren Disparität	95
3.42 Abhängigkeit zwischen der Entfernung zum Stereokamerasystem und der Disparität	97
3.43 Fehler bei der Bestimmung der Entfernung aufgrund einer fehlerhaften Disparität	99
3.44 Disparität eines Versuches ohne Marker	100
3.45 Höhenfeld des Disparitätsbildes als Rückprojektion in das Weltkoordinatensystem	101
3.46 Falsch positive Detektionen durch fehlende Hintergrundsubtraktion	102
3.47 Maximaler Hintergrundabstand	103
3.48 Vordergrunderosion durch perspektivisches Höhenfeld	103
3.49 Hintergrundsubtraktion und ihre manuelle Anpassung	104
3.50 Schritte der markerlosen Personendetektion	105
3.51 Ausschnitte der Schritte der markerlosen Personendetektion	106
3.52 Perspektivisches Höhenfeld	107
3.53 Interpolation innerhalb der Disparitätsmatrix	107
3.54 Auswahl von Isolinien	108
3.55 Fälle bei der Zuordnung neuer Ellipsen zu bestehenden Ellipsenpyramiden	110
3.56 Fall 4 bei der Zuordnung neuer Ellipsen zu bestehenden Ellipsenpyramiden	111
3.57 Standardabweichung von Halbachsen in einer Schnittebene und einem Winkelintervall	112
3.58 Halbachsen für PES(10 cm)	114
3.59 Kombinationen von Halbachsen im PES	114
3.60 Durchschnittliche PES hinterlegt mit durchschnittlicher Person	115
3.61 Durchschnittliche PES in sechs äquidistanten Blickwinkel-Intervallen	117
3.62 Reziproke Steigung der Mittelachsen von durchschnittlichen PES	117
3.63 Halbachsen der Ebenen innerhalb des PES in Abhängigkeit vom Blickwinkeln	118
3.64 Positionsfehler bei Kasernen-Versuchen	121
3.65 Ungleichmäßiger Verlauf der Personengröße	124
3.66 Laufwege eine Treppe herunterlaufender Personen	124
3.67 Verteilung der Personengrößen laut Befragung und automatischer Erfassung	125

3.68	Änderung der Personengröße entlang des Laufweges	126
3.69	Ähnliche Struktur wie der Hermes-Marker durch Schnüre der Kopfbedeckung	129
3.70	Entfernungsabhängig eingefärbte Überkopfaufnahme einer Stereokamera	130
3.71	3D Laufwege zweier überlappender Kameras	130
3.72	Laufwege eingefärbt entsprechend des Kameraabstandes	131
3.73	Laufwege eingefärbt entsprechend der Shirtfarbe	131
3.74	Parameteranpassung des Farbmarkers	134
3.75	Skizze zum Fehler der Lage des Farbmarkers	135
3.76	Perspektivisch abgebildeter Bereich farbiger Mützen	137
3.77	Fehler des Farbmarkers abhängig vom Betrachtungswinkel und der Kameraentfernung	137
3.78	Virtuelle Person mit Farbmarker	138
3.79	Trajektorie und Geschwindigkeit einer über Farbmarker erkannten virtuellen Person	139
3.80	Erkennung einer Person in einer virtuellen Szene	143
3.81	Ausschnitt von Erkennungsschritten einer Person in einer virtuellen Szene	144
3.82	Disparitätsmatrix mit Höhenlinien in Abhängigkeit von der Vergleichsfenstergröße	145
3.83	Laufweg einer virtuellen Person in Abhängigkeit von der Vergleichsfenstergröße	145
3.84	Markerlose Personendetektion in einer T-Kreuzung	149
3.85	Ausschnitt der markerlosen Personendetektion	150
3.86	Glattheit der Laufwege aufgrund des Einsatzes von Markern	151
3.87	Variierende Personendichte virtueller Personen	153
3.88	Isolinien und Pyramiden bei unterschiedlichen Personendichten	154
3.89	Markerlose Erkennung bei Schrägsicht der Kamera	155
3.90	Position bei Verfahren mit und ohne Marker	157
3.91	Trajektorie und Geschwindigkeit bei Verfahren mit und ohne Marker	158
3.92	Vergleich Laufweg mit und ohne Marker	159
3.93	Geschwindigkeit bei Verfahren mit und ohne Marker	161
3.94	Zähltor und Bild eines Erfassungssystems einer Feldstudie	164
3.95	Erfassungssysteme am Zähltor einer Feldstudie	164
3.96	Skizze des Zähltores mit Erfassungssystemen	165
3.97	Zählraten der kommerziellen Erfassungssysteme	165
3.98	Bilder der Stereokamera mit Laufwegen einer Feldstudie	167
3.99	Disparität des Hintergrundes	168
3.100	Versuch der Technischen Universität Dänemarks	172
3.101	Versuch der Firma SL-Rasch	172
3.102	Dichte-Profil-Änderung bei steigender Falsch-Negativ-Rate	175

3.103 Dichte-Verteilung bei steigender Falsch-Negativ-Rate	176
3.104 Personenabstand in Abhängigkeit von der Berücksichtigung der Personengröße	178
3.105 Fehlereinfluss auf Linienbildung	179
4.1 Kameraeinstellungen zu Versuchsreihe 1, 2, 4 und 5	184
4.2 Kameraeinstellungen zu Versuchsreihe 3 und 7	185
4.3 Kameraeinstellungen zu Versuchsreihe 6 und 8	186
4.4 Maximaler Bildwinkel ohne Verdeckung	187
4.5 Einschränkung des Bildwinkels an Treppen	188
4.6 Abdeckung durch Kamera-Grid	189
4.7 Versetzte Kameraanordnung verringert maximalen Betrachtungswinkel	189
4.8 Experimente auf der Tribüne der ESPRIT Arena	190
4.9 Experimente an der Außentreppe der ESPRIT Arena	191
4.10 Bestimmung der Pixelabdeckung und Brennweite	192
4.11 Bewegungsunschärfe	193
4.12 Ausgefrante Kopfform durch Zeilensprungverfahren	194
4.13 Kabelwege zur Bildübertragung und Synchronisation	196
4.15 Versuchsteilnehmer	198
4.14 Kameraaufhängung und Versuchsaufzeichnung	199
4.16 Arbeitsplatz zur Versuchserfassung	201
4.17 Glättung von Trajektorien aufgrund geschwindigkeitsabhängiger Schwankung	203
4.18 Voronoi-Diagramm eines Korridor-Experiments	204
4.19 Geschwindigkeits-, Dichte- und Fluss-Profil eines Korridors mit Einschnürung	205
4.20 Raum-Zeit-Diagramme von Bewegungen von Fußgängern und Fahrradfahrern	207
4.21 Aufbauskizze und Aufnahme von Engstellenversuchen	208
4.22 Abhängigkeit des Flusses innerhalb einer Engstelle von ihrer Breite und Länge	208
4.23 Laufwege durch Engstellen mit unterschiedlicher Breite und Länge	209
4.24 Dichte-Profil von Engstellen unterschiedlicher Breite	209
4.25 Fluss-Profil einer Engstelle unterschiedlicher Länge	210
4.26 Geschwindigkeit, Dichte und spezifischer Fluss im Querschnitt einer Engstelle	211
4.27 Aufbauskizze und Aufnahme von T-Kreuzungs-Versuchen	211
4.28 Schrägsicht auf ein T-Kreuzungs-Versuch	211
4.29 Laufwege zusammenfließender Personenströme für unterschiedliche Zuflussbreiten . .	212
4.30 Fundamentaldiagramm unterschiedlicher Messbereiche einer T-Kreuzung	212
4.31 Vergleich von Experiment und Simulation anhand der T-Kreuzung	213

4.32	Dichte-, Geschwindigkeits- und Fluss-Profil einer T-Kreuzung	214
4.33	Schema des Evakuierungsassistenten im Projekt Hermes	215
4.34	Vergleich von Experiment und Simulation anhand der Räumung einer Tribüne	216
4.35	Kommunikationsmodul und Vorhersageergebnis des Evakuierungsassistenten	216
5.1	Messung in einer Schrägsicht eines Fahrradexperimentes	220
5.2	Glättung des Laufweges durch Einhüllende	222
5.3	Tiere als selbstgetriebene Teilchen	222
D.1	Logo der Software PeTrack	290
D.2	Aufrufgraph vom Konstruktor der Klasse Petrack	291
D.3	Grafische Benutzeroberfläche der Software PeTrack	293
D.4	Eingabefenster und Menüs der Menüleiste von PeTrack	293
D.5	Registerkarten der grafischen Benutzeroberfläche von PeTrack	294

Tabellenverzeichnis

2.1	Laborversuche anderer Forschergruppen	16
2.2	Tabellarischer Vergleich aktueller Verfahren zur Personendetektion	23
2.3	Fähigkeiten verschiedener Sensorarten zur Personendetektion	42
2.4	Bewertung vorgestellter Techniken	44
3.1	Eigenschaften verwendeter Marker	60
3.2	Metrische Größen der in Abb. 3.40 markierten Punkte	94
3.3	Erkennungsfehler in Abhängigkeit von der Mindestgröße der PES	115
3.4	Fehlerraten der Detektion durch Marker mit Strukturelementen	128
3.5	Dichteabhängige Fehlerraten bei Nutzung des Hermes-Markers	129
3.6	Abstandsfehler in Abhängigkeit von Verfahren und Vergleichsfenstergröße	146
3.7	Erkennungsfehler in Abhängigkeit von Verfahren zur Disparitätsbestimmung	147
3.8	Fehler der markerlosen Personenerkennung	150
3.9	Dichteabhängige Erkennungsrate	152
3.10	Versuchsabhängiger Zeitbedarf der Schritte zur Extraktion der Trajektorien	170
3.11	Fehlerraten aller Erkennungsmethoden	180
4.1	Durchgeführte Experimente	182
4.2	Eingesetzte Kameras	194
4.3	Auf den Experimenten beruhende Veröffentlichungen	206

1

Einleitung

1.1 Motivation

Die Dynamik von Fußgängerströmen beeinflusst viele Situationen im alltäglichen Leben und offenbart dabei interessante Selbstorganisationsphänomene und kollektive Effekte. Obwohl für uns die Bewegung als Fußgänger in Menschenmengen alltäglich ist, wurde bisher auf diesem Gebiet wenig Grundlagenforschung betrieben.

Dabei würde ein tiefergehendes Verständnis eine Optimierung von Gebäuden und Anlagen für den Fußgängerverkehr sowohl in sicherheitstechnischer als auch ökonomischer Hinsicht ermöglichen und den Komfort für den Nutzer steigern helfen. Bei Gebäuden für große Personenzahlen (Arenen, Theater etc.; siehe Abb. 1.1) und Verkehrsinfrastrukturen (Bahnhöfe, Flughäfen etc.) kommt den Verkehrswegen, insbesondere den Flucht- und Rettungswegen als sicherheitsrelevanten Infrastrukturen, eine zentrale Bedeutung zu. Aber auch im Hinblick auf wachsende Städte [1, 2] und die damit einhergehende verstärkte Nutzung öffentlicher Verkehrsmittel sowie den häufiger und größer werdenden öffentlichen Veranstaltungen (z. B. Public Viewing) ist das Verständnis über die Dynamik von Menschenmassen von großer Bedeutung.



Abb. 1.1: Konzerte mit großen Menschenmengen: in der multifunktionalen ESPRIT Arena in Düsseldorf in einer komplexen Gebäudestruktur und „Rock am Ring“ auf offenem Gelände; Quelle Bild „Rock am Ring“: Thomas Hußmann, Feuerwehr Düsseldorf



Abb. 1.2: Heilige Moschee in Mekka während des Haddsch; Bild aus [3]

Generell ist die Tendenz zu größeren Veranstaltungen zu erkennen. Dieser Anstieg der Personenzahlen ist auch bei religiösen Festen zu verzeichnen, z. B. bei dem Haddsch (Islam, siehe Abb. 1.2 und [3]) oder Kumbh Mela (Hinduismus, siehe [4]), bei denen es in der Vergangenheit leider häufig zu Unglücken mit Personenschäden kam.

Das Video Nr. 40 auf Seite 286 zeigt verschiedene Orte von großen Fußgängerströmen. Video Nr. 37 ist bei der Loveparade in Duisburg kurz vor dem tragischen Unglück mit 21 Toten entstanden und gibt eine Situation wieder, zu der es bei guter Planung und damit richtiger Dimensionierung der Fußgängeranlagen nie kommen sollte.

Typische Fragen sind dabei: Ist die Breite einer Tür oder eines Korridors ausreichend, um eine gewisse Anzahl an Personen in einer gegebenen Zeit passieren zu lassen oder zu evakuieren? Wie lange dauert die Räumung eines Gebäudes?

Die verfügbaren Methoden und Werkzeuge, um solche Fragen zu beantworten, kann in folgende Gruppen aufgeteilt werden: gesetzliche Regelungen [5, 6], Handbücher [7–9] und Computersimulationen [10–13].

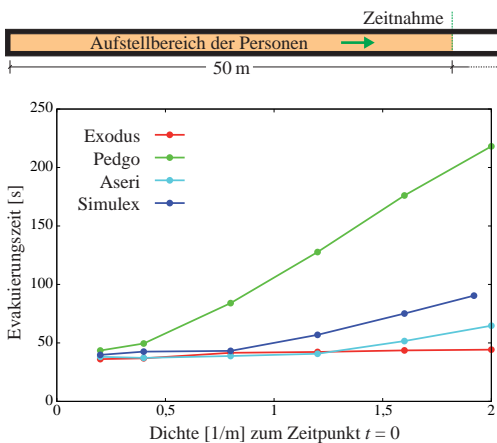


Abb. 1.3: Stark variierende simulierte Evakuierungszeiten eines Korridors mit unterschiedlicher Personendichte; Bild basiert auf Abbildung in [14]

Gesetzliche Regelungen basieren auf Vorschriften mit statischen Regeln, die u. a. von der Belegung oder der Größe der Verkaufsfläche eines Gebäudes abhängen: beispielsweise die minimale Türbreite in Abhängigkeit von der Personenzahl in einem Raum oder die maximale Länge eines Fluchtweges. Jedoch können statische Regeln nicht die Dynamik eines Evakuierungsprozesses abbilden, so dass Methoden mit der Möglichkeit zur Abbildung chronologischer Prozesse nötig sind, um die Entwicklung über die Zeit zu berücksichtigen.

Handbücher nutzen dafür makroskopische Modelle von Personenströmen und erlauben somit eine zeitliche und räumliche Beschreibung des Evakuierungsprozesses. Diese Modelle können vorhersagen, wo und wann es zu Stauungen kommt. Trotzdem bieten sie nur eine grobe Beschreibung, da der Personenstrom

als eine Einheit mit konstanter Dichte beschrieben wird.

Mikroskopische Modelle beschreiben hingegen die Bewegung jedes Einzelnen und bieten somit die Möglichkeit die individuelle Dynamik abzubilden. Nur dadurch sind komplexe Bewegungen von Menschenmassen, wie z. B. kreuzende oder gegenläufige Ströme, abbildbar. Ungeachtet der Diffe-

renziertheit bzw. Komplexität eines Modells sollte für sicherheitsrelevante Anwendungen das Ziel die quantitativ richtige Beschreibung eines Personenstroms sein.

Der Stand der Wissenschaft und Technik ist in diesem Bereich jedoch ungenügend. Dies wird unter anderem an widersprüchlichen gesetzlichen Regelungen zur Gestaltung von Rettungswegen deutlich (siehe Abschnitt 4.4.1) und zeigt sich auch in stark unterschiedlichen Ergebnissen kommerzieller Simulationsprogramme zur Evakuierung auch einfachster Szenarien [15, 16]. Ein Beispiel zeigt Abb. 1.3, die die großen Unterschiede in der berechneten Evakuierungszeit für das einfache Szenario eines zu entleerenden schmalen Korridors verdeutlicht. Die Evakuierungszeit variiert für die besonders sicherheitsrelevanten hohen Dichten um einen Faktor von 4.

Bereits grundlegende Fragen werden in der Literatur widersprüchlich diskutiert [17]: z. B. wie der maximal mögliche Fluss durch eine Engstelle von deren Breite abhängt [18, 19], ab welcher Dichte eine Stauung auftritt [20] oder ob sich der Fluss eines uni- und eines bidirektionalen Stroms unterscheidet [21].

Ein zentraler Grund für das ungenügende Wissen ist die rudimentäre und widersprüchliche empirische Datenbasis [17]. Abb. 1.4 zeigt diese widersprüchlichen Daten und die damit verbundenen uneinheitlichen Annahmen in Handbüchern (siehe Abschnitt 4.4.1). Das sogenannte Fundamentaldiagramm der Fußgängerdynamik, hier in Form eines Dichte-Geschwindigkeits-Diagramms, zeigt die große Streuung in den experimentellen Daten sowie die geringe Datenbasis für hohe Dichten. Die freie Gehgeschwindigkeit bei geringen Dichten nahe $\rho = 0$ wird in den Handbüchern mit 1,0 m/s bis 1,4 m/s angegeben und die Stoppdichte bei der Geschwindigkeit $v = 0$ variiert zwischen $3,8/\text{m}^2$ und $10/\text{m}^2$. Die Dokumentation der Erhebungen, aus der die Unterschiede hätten ersichtlich sein können, ist leider oft ungenau (u. a. Randbedingungen, Messverfahren).

Es fehlt somit an belastbaren Daten für das Verständnis der Fußgängerdynamik. Daher haben wir fast 600 das gesamte Dichtespektrum umfassende Laborexperimente durchgeführt (siehe Abschnitt 4.1) und dabei die individuellen Laufwege erfasst (siehe Kapitel 3). Mit Hilfe der Laufwege lassen sich die relevanten Größen eines Personenstroms wie Geschwindigkeit, Dichte und Fluss bestimmen (siehe Abschnitt 4.2). Sie ermöglichen die obigen Fragestellungen gezielt zu untersuchen und so das Verständnis der Fußgängerdynamik zu erhöhen (siehe Abschnitt 4.4). Modelle können auf dieser Basis die beobachteten Zusammenhänge abbilden, mit den erhobenen Daten kalibriert und auch validiert werden (siehe kursive Veröffentlichungen in Tab. 4.3). Mikroskopische Modelle benötigen dabei auch mikroskopische Daten. Durch die Simulationen können schließlich komplexere Fragestellungen wie der Personenfluß durch ein großes Stadion beantwortet werden.

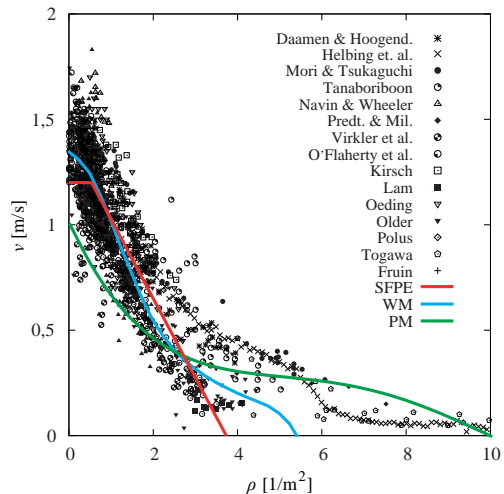


Abb. 1.4: Fundamentaldiagramm der Fußgängerdynamik als Dichte-Geschwindigkeits-Diagramm; Datenpunkte stehen für empirische Erhebungen und Linien für Annahmen in Handbüchern; siehe [17, 22]

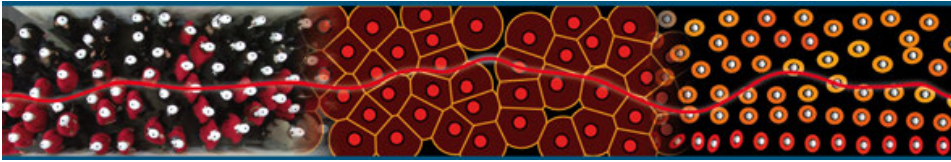


Abb. 1.5: Illustration zeigt von links nach rechts den Weg vom Experiment, über die Erfassung der Laufwege, deren Analyse bis hin zu dem damit entwickelten Modell zur Simulation der Personenbewegung

1.2 Ziel und Aufbau der Arbeit

Diese Arbeit hat zum Ziel, die Methodik der Erhebung der empirischen Daten bei unseren Experimenten zum Strömungsverhalten von Fußgängern zu beschreiben. Das Resultat sind idealerweise Laufwege jeder einzelnen Person.

Bei unseren ersten Versuchsreihen (siehe Tab. 4.1) mit einer geringen Anzahl an Durchläufen und Personen wurden mit großem zeitlichen Aufwand Handauswertungen durchgeführt, die jedoch nur die Erfassung von Durchgangszeiten der Probanden an einigen Stellen erlaubten. Es folgten halbautomatische Ansätze, bei denen die Personen manuell lokalisiert und mittels Tracking-Verfahren verfolgt wurden. Dadurch waren zum ersten Mal individuelle Laufwege verfügbar.

Aber erst mit voll automatischen Verfahren und mit Markern versehenen Personen war eine zeitlich fein aufgelöste und mit wenig Fehlern behaftete Auswertung von vielen Versuchen mit einer großen Anzahl an Personen effektiv möglich. Die Entwicklung der Methoden zur automatischen Erfassung soll Schwerpunkt dieser Arbeit sein. Dabei muss die automatische Auswertung auch für hohe Personendichten funktionieren. Als Resultat der Implementation der vorgestellten Verfahren entstand die freie Software *PeTrack*.

Da für das Studium von Phänomenen, die nur unter realen Bedingungen auftreten (z.B. Verhalten von frustrierten oder euphorischen Fußballfans), neben den Laborversuchen auch Feldstudien durchgeführt werden sollen, ist ebenfalls ein Verfahren zur Erfassung ohne Markierung der Personen entwickelt worden. Diese markerlose Detektion muss aufgrund der besonders interessierenden Dynamik in Stausituationen auch eine Erfassung individueller Laufwege bei hohen Personendichten ermöglichen.

Die Arbeit des Autors bestand bei den Experimenten zum einen in der Erfassung der Versuche, von der Konzeptionierung über die technische Umsetzung bis hin zur Durchführung, und zum anderen in der anschließenden automatischen Auswertung der erfassten Aufzeichnungen. Hierzu wurden verschiedene Verfahren entwickelt, die auf die Rahmenbedingungen der Versuche und der Örtlichkeit, in der diese durchgeführt wurden (u. a. Personendichte, Abdeckung bzw. Entfernung zwischen Erfassungstechnik und Probanden, Geometrie wie z. B. Treppen, Laborbedingung oder Feldstudie), abgestimmt waren. Die Laufwege jeder einzelnen Person bzw. die Trajektorien sollen dabei möglichst exakt sein, um Größen wie Geschwindigkeit, Fluss und Dichte an jeder Stelle bestimmen zu können und damit die Analyse auch auf einer mikroskopischen Skala zu ermöglichen.

Besondere Herausforderungen bestanden dabei u. a. in der Auslegung der Aufzeichnungstechnik (siehe Abschnitt 4.1.1), der möglichst präzisen Bestimmung der metrischen Position einer Person, die die optische (siehe Abschnitt 3.1) und perspektivische Verzerrung (siehe Abschnitt 3.4) der Objekte berücksichtigt, und der Auslegung sowie Erkennung der Marker (siehe Abschnitt 3.2), die stark von den Rahmenbedingungen abhängig waren. Da jeder Versuchsaufbau im gesamten Dichtespektrum untersucht wurde, musste die Funktionalität der Verfahren immer auch bei hohen Personendichten

gewährleistet sein, was für die Erkennung die größte Herausforderung mit sich brachte. Für die markerlose Detektionsmethode bestand die Schwierigkeit neben der eigentlichen Methodenentwicklung in der Wahl und Abstimmung der Vielzahl an Parametern, u. a. bei der Disparitätsbestimmung (siehe Abschnitt 3.4.2), den Strukturmerkmalen (siehe Abschnitt 3.5.2 bis 3.5.4) und der Kategorisierung der Kandidaten (siehe Abschnitt 3.5.5), im Prozess der Identifikation, so dass eine stabile mit wenig Fehlern behaftete Positionsbestimmung (siehe Abschnitt 3.6.6) erfolgen kann.

Da die Verfahren zur automatischen Extraktion von Laufwegen Schwerpunkt dieser Arbeit sind, werden in Kapitel 2 hierfür der Stand der Technik und existierende Verfahren vorgestellt, bevor in Kapitel 3 die eigenen Entwicklungen beschrieben werden. Für den Leser, der den gesamten kontinuierlichen Prozess der Datenerhebung nachvollziehen möchte, sei folgende Lesereihenfolge empfohlen: Kapitel 4 bis Abschnitt 4.1, Kapitel 3 und anschließend Abschnitt 4.2 bis 4.4.

Für eine umfassende Darstellung des Themas sind neben einem Ausblick in Kapitel 5 und der Literaturliste im Anhang eine Liste von Zeitschriften und Konferenzen zu diesem Themengebiet, Verweise auf vornehmlich eigene Videos, eine Beschreibung der Implementation der vorgestellten Verfahren, eine Liste von kommerziellen Produkten zur Erfassung von Personen sowie ein Glossar aufgeführt.

2

Existierende Verfahren zur Lokalisierung von Personen

In diesem Kapitel soll ein Überblick über Verfahren, Techniken und Einsatzgebiete der Detektion von Personen gegeben werden. Die Beschäftigung mit automatischen Verfahren zur Erfassung der Laufwege von Personen begann 2005 nach der Durchführung von Versuchsreihe Nr. 2 aus Tab. 4.1, da die manuelle Ermittlung sehr aufwendig und trotzdem ungenau war. Vor einer ausführlichen Literaturrecherche wurden zuerst einfache Methoden getestet, ob diese bereits erlauben, Personen zuverlässig zu detektieren. Da die Aufzeichnung aller Versuche mit Kameras durchgeführt wurde, wurden direkte Merkmale in den Bildern gesucht, die die Position einer Person beschreiben.

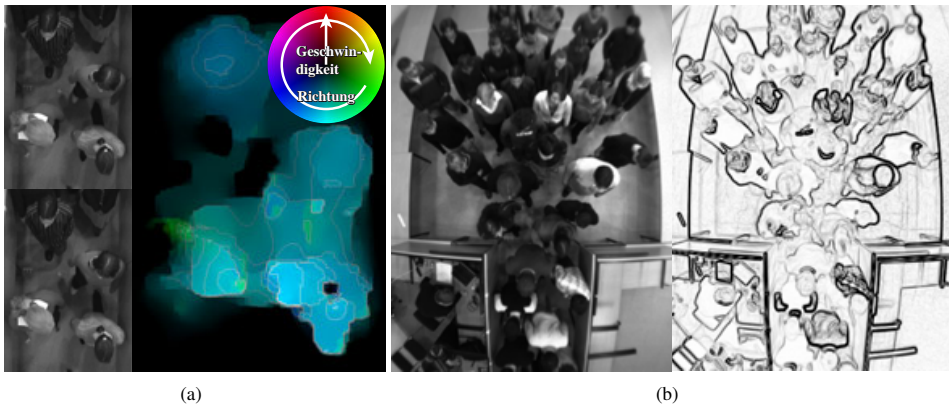


Abb. 2.1: Engstellenexperiment aus Versuchsreihe Nr. 2 mit Kameraperspektiven (a) innerhalb der Engstelle und dem optischen Fluss zwischen zwei Folgebildern und (b) vor der Engstelle und dem Gradientenbild der Helligkeit auf der Suche nach Merkmalen zur Personendetektion

Abb. 2.1 a zeigt zwei aufeinanderfolgende Videobilder eines Versuches aus Versuchsreihe Nr. 2 in einer Engstelle und den optischen Fluss [23, 24] jedes Pixels zwischen diesen Bildern, der die Bewegung in der Bildebene beschreibt. Der optische Fluss ist ein Vektorfeld an den Bildpunkten, das die

Bewegung der Bildpunkte in der Bildebene zwischen zwei Bildern angibt. Der Begriff des „optischen Flusses“ wird neben weiteren Fachbegriffen dieser Arbeit ebenfalls im Glossar erläutert (Anhang G). Zur Bestimmung des optischen Flusses ist ein Verfahren von Michael J. Black verwandt worden [26]. Die erste Annahme, dass Maxima im optischen Fluss oder Regionen großer Pixelbewegungen Köpfe von Personen eindeutig identifizieren lassen, weil sie näher an den überkopf aufgehängenen Kameras liegen, trifft nicht zu, da wie hier durch Drehung des Kopfes und einen schnell bewegenden Zopf, oder beim Gehen bewegende Beine und Füße, auch andere Bereiche große optische Flüsse aufweisen. Zudem kommt es in Experimenten insbesondere an Engstellen oftmals zu Stauungen und damit zu sich kaum bis gar nicht bewegenden Personen, so dass der optische Fluss nahe Null liegt und nicht alleinig zur Identifizierung herangezogen werden kann.

Zur Erkennung von Objekten nutzen Menschen neben Bewegungsmerkmalen auch Struktur- bzw. Konturmerkmale [27, 28]. Abb. 2.1 b zeigt ein Bild des zuvor betrachteten Engstellenversuchs, jedoch vor und innerhalb der Engstelle und das entsprechende Gradientenbild der Helligkeit [30], das Kanten im Grauwertbild herausstellt (siehe Video Nr. 29). Es ist ersichtlich, dass eine stabile Detektion auch hiermit nicht direkt möglich ist, da Personen sich unterschiedlich vom Hintergrund abheben und auch die Textur der Kleidung starken Einfluss auf das Gradientenbild hat. Trotzdem sind, wie später beschrieben, Histogramme der Gradienten eines der erfolgreichsten Verfahren (siehe Abschnitt 2.3.1.3).

Auch ein Differenzbild [30] zweier aufeinanderfolgender Bilder wäre zur Personenerkennung denkbar (siehe Abb. 3.20), aber auch auf dieses Merkmal hat die Unterscheidbarkeit der Personen vom Hintergrund großen Einfluss und zudem ist bei einer oftmals bei Versuchen vorkommenden Staubildung keine Identifizierung möglich, da sich Folgebilder kaum voneinander unterscheiden.

Nachdem kein einfaches direktes Verfahren gefunden wurde, wurde eine ausführliche Literaturrecherche nötig (Liste möglicher Quellen siehe Anhang B), die eine Vielzahl von Ideen aufzeigte, aber deren Resultate von der angestrebten Genauigkeit und Fehlerrate auch bei hohen Personendichten noch weit entfernt waren. Daraufhin wurde in einem nächsten Schritt die Personenerkennung bei Laborversuchen, in denen alle Möglichkeiten der Einflussnahme bestehen, mittels Markierung jeder einzelnen Person durchgeführt (siehe Abschnitt 3.2). Erst 2011 wurde ein Verfahren entwickelt, das eine Erfassung von Laufwegen in Feldstudien ermöglicht [31, 32]. Dieses diente zudem der Verbesserung der Detektion mit Markern (siehe Abschnitt 3.5).

Dieses Kapitel gibt einen Überblick über existierende Verfahren zur Lokalisierung von Personen. Dabei ist zuerst eine Abgrenzung zu ähnlichen Gebieten nötig.

2.1 Abgrenzung zu ähnlichen Gebieten

Spricht man von der Identifikation oder Erfassung einer Person oder deren Bewegung, so sind unterschiedliche Detaillierungsgrade denkbar. Ist es für die eine Anwendung ausreichend zu wissen, ob überhaupt eine Person oder ein Objekt im Bild oder in einem Raum ist oder in welche Richtung sich eine ganze Menschenmenge und nicht ein einzelnes Individuum bewegt, so zielen andere Verfahren darauf ab, eine Person z. B. über biometrische Daten zu identifizieren oder wiederzuerkennen.

Mit steigendem Detaillierungsgrad können folgende Stufen der Erfassung von Personen und ihrer Bewegungen unterschieden werden (in Klammern Beispiele typischer Fragestellungen bzw. resultierender Daten):

1. Menschenmassen in ihrer Gesamtheit (Anwesenheit von Personen in bestimmten Bereichen),
2. Dynamik innerhalb der Menschenmasse ohne Unterscheidung von Einzelpersonen (Geschwindigkeiten, Flussrichtung, siehe Abb. 2.2),
3. Einzelpersonen ohne genaue Ortsangabe (großräumige Laufwege ohne detaillierte Positionen),
4. Einzelpersonen (Trajektorien),
5. identifizierte Einzelpersonen (individualisierte Trajektorien),
6. Körper (Kopfausrichtung, Bewegung von Gliedmaßen des Körpers, z. B. Schrittfolge, siehe Abb. 2.3),
7. Körperdetails (Bewegung von Fingern, Blickrichtung).

In [33] wird folgende Unterteilung der Verfahren mit steigendem Informationsgehalt vorgeschlagen: Vorkommen, Zählung, Lokalisierung, Laufweg und Identifizierung.

Dabei sind grundlegende Aufgaben je nach Detaillierungsgrad die Erfassung von Personen(gruppen), die Verfolgung von Personen(gruppen), die Erkennung von individuellen Eigenschaften und die Aktivitätsanalyse.

Einige Anwendungsgebiete und -beispiele sind:

- kollisionsfreie Navigation im öffentlichen Raum [34, 35],
- Leitung und Überwachung von Menschenmengen [36, 37] (*crowd management*) z. B. bei der Durchführung von Großveranstaltungen, u. a. durch eine großskalige Erfassung von Bewegungsmustern oder der übergeordneten Bewegungsrichtung der Menge (*visual surveillance*) (siehe unten),
- Bewegungserfassung (*motion capturing*) und -analyse des Körpers im Bereich des Sports, der Medizin oder des Films auf unterschiedlichen Skalen (siehe unten),
- Aktivitätsanalyse bzw. Analyse des Verhaltens von Personen (*behaviour analysis*) oder speziell die Erkennung von Gefahren (*event detection*) z. B. abnormale Bewegungen einer Person beim Kämpfen, einer Rangelei oder beim Fallen [38, 39],
- Interaktion mit virtuellen Umgebungen mittels des ganzen Körpers oder durch Gesten der Gliedmaßen oder Finger z. B. in Virtual Reality (VR)-Umgebungen oder in Spielen [40],
- Identifizierung oder Wiedererkennung von Personen [41–43] z. B. mittels Gesichtserkennung (*automatic face recognition*) [44–46] oder Ganganalyse (*automatic gait recognition*) [47] zur Zutrittskontrolle oder der Verfolgung [41, 48],
- Verstehen von Szenen für automatische Handlungen z. B. eines Roboters bei der Bergung von Personen [49] oder der Erkennung von Alter und Geschlecht [50, 51] zur individuellen Behandlung,
- Frequentierung oder Überwachung von Bereichen z. B. in Geschäften oder zutrittsbeschränkten Gebieten [52–54],

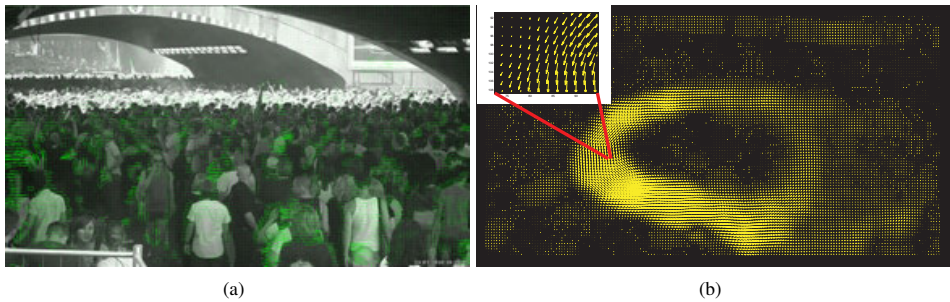
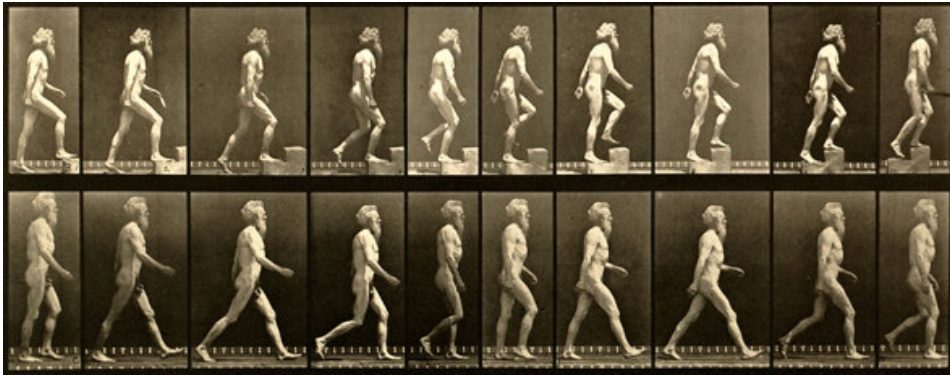
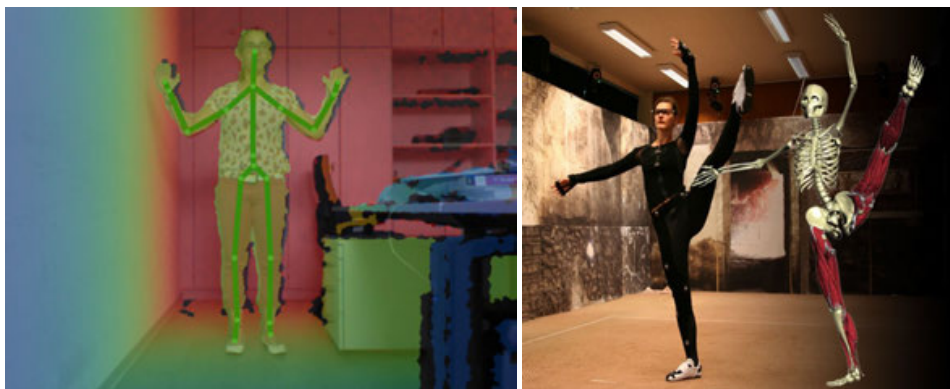


Abb. 2.2: Optischer Fluss, in der Anwendung ggf. als Strömungsfeld (*flow field*) bezeichnet, für eine Auswahl von Bildpunkten an Kreuzungspunkten eines kartesischen Gitters zur Untersuchung von Bewegungen in Menschenmassen: (a) Teilnehmer der Loveparade 2010 in Duisburg auf dem Weg zum Veranstaltungsgelände mit überlagerten grünen Vektoren zur Darstellung des optischen Flusses; (b) Pilger in Mekka, die in Bereichen großen Flusses bzw. langer gelber Vektoren die Kabba umkreisen (vgl. Abb. 1.2); Bilder aus (a) [59] und (b) [60]



(a)



(b)

(c)

Abb. 2.3: Bewegungsanalyse bzw. -erfassung des gesamten Körpers durch (a) Chronofotografie, (b) Microsoft Kinect und (c) Erfassungssystem im MIRALab der Universität Genf; Bild aus (a) [61], (c) [62]

- Erfassung von detaillierten Laufwegen zur mikroskopischen Analyse der Dynamik, z. B. zur Modellbildung, -kalibrierung und -validierung [55–58].

Abb. 2.2 zeigt Beispiele einer Erfassung von Merkmalen im optischen Fluss zur Analyse von Bewegungen innerhalb von Menschenmassen. Der optische Fluss wird in [59] verwendet, um anhand von Bewegungsmustern in einer Menschenmasse das Eintreten von Ereignissen aufzuspüren. In [60] wird der optische Fluss, wie in Abb. 2.2 b zu sehen, u. a. zur Segmentierung von Menschenmassen in Bereiche unterschiedlicher Aktivitäten genutzt. Das Verhalten von Personen, Personengruppen oder Menschenmengen (z. B. Wellen unkontrollierbarer Bewegung) innerhalb von Szenen zu detektieren, in denen Einzelpersonen zumindest zeitweise komplett zu sehen sind, ist in [52, 63] beschrieben. Die Bewertung von Gruppenverhalten, insbesondere die Detektion von abnormalem Verhalten, oder ungewöhnliche Bewegungen in Menschenmassen wird in [38, 64–71] behandelt. Dabei werden in [59, 72–83] keine Individuen separiert, sondern Bewegungsmuster von Gruppen analysiert. Eine häufige Anwendung ist die Dichteschätzung, für die nicht exakt jede Person lokalisiert werden muss, sondern u. a. über Lernalgorithmen Merkmale bzw. Texturen in einer Menschenmenge einer Dichte zugeordnet werden [84–88]. Mit Bestimmung der Geschwindigkeit über den optischen Fluss [89] ist es zusammen mit einer ggf. geschätzten Dichte möglich, den Personenfluss zu berechnen, ohne Einzelpersonen detektieren zu müssen.

In Abb. 2.3 sind Beispiele eines Studiums bzw. der Erfassung von Bewegungsabläufen des Körpers dargestellt. Die obere Bilderreihe gibt den Ablauf beim Treppe steigen und Gehen mit und von Eadward Muybridge, einem bedeutenden frühen Vertreter der Serienfotografie bzw. Chronofotographie [90], wieder. Die untere Reihe der Abb. 2.3 stellt zwei Anwendungen der Bewegungserfassung des Körpers (*motion capturing*) [91] dar. In b wird das mittels eines Kinect-Sensors von Microsoft [92] (siehe Abschnitt 2.3.7) ermittelte Tiefenbild und das Ergebnis der erkannten Stellung einer Person als Überlagerung auf das reale Bild gezeigt, dessen Ergebnisse zur Interaktion mit dem Spielgeschehen genutzt wird. Beim Film wird die detaillierte Bewegungserfassung des Körpers u. a. dazu benutzt, um virtuelle Charaktere sich realistisch bewegen zu lassen, was in c mittels eines Erfassungssystems im MIRALab der Universität Genf [62] durchgeführt wurde.

2.2 Verfahren zur Erfassung der Position einer Person

Für die mikroskopische Simulation der Dynamik in einer Menschenmenge, in der die Bewegungen jeder einzelnen Person simuliert werden, sind die oben beschriebenen erfassten Merkmale entweder zu ungenau (z. B. Bewegungsmuster innerhalb einer Menschenmenge) oder beinhalten Details (z. B. Gesichtsmerkmale), die die Dynamik nicht beeinflussen oder in die Modellierung derzeit noch nicht einfließen. Der benötigte Detaillierungsgrad liegt für die vorausgehende Analyse als Basis der Modellentwicklung und der Kalibrierung und Validierung der berechneten Ergebnisse bei der Erfassung von exakten Laufwegen jeder einzelnen Person.

Eine Bewertung der erfassten Laufwege muss nicht direkt vom System, wie z. B. bei einem Überwachungssystem, durchgeführt werden.

Es ist jedoch denkbar, dass einige Merkmale (z. B. Gang, Blickrichtung), die bisher von mir noch nicht erfasst wurden, relevant für die Dynamik sind und damit in zukünftige Modelle einfließen. Bei den hohen Dichten, die teilweise in den von uns durchgeführten Experimenten vorherrschten, und den verwendeten Kamera-Techniken sind jedoch außer den Oberkörpern und den mitgeführten Markern keine weiteren Details erkennbar. Die Kopfausrichtung als etwas schwächere Information wie die

Blickrichtung kann bereits mit zwei der im folgenden Kapitel vorgestellten Marker direkt ausgelesen werden. Für die Zukunft ist geplant, die Blickerfassung bzw. Okulographie (*eye tracking*) mittels Eye-Tracker für einzelne Probanden durchzuführen.

Es folgt im Weiteren eine Übersicht über existierende Verfahren zur Ermittlung der Positionen von Personen. Da das Tracking der einmal erkannten Personen für unsere Anwendung, in der wir die Kameras überkopf aufhängen können und es kaum zu Verdeckungen kommt, stabil mit einem etablierten Verfahren funktioniert (siehe Abschnitt 3.3), wird darauf bei der Vorstellung von vorangegangenen Arbeiten nicht detailliert eingegangen. Im Kontext der Personendetektion wird dies jedoch in anderen Arbeiten ausführlich behandelt, da es bei einem großen Bereich einsehenden schrägen Kameraperspektiven (z. B. Überwachungskameras) häufig zu Verdeckungen kommt. Zur Gliederung der Verfahren zur Personendetektion könnten diese u. a. nach folgenden Kriterien unterteilt werden:

- verwendete Merkmale zur Detektion (z. B. Form oder Farbe bzw. objekt- oder pixelbasiert),
- maximale Personendichte, für die das jeweilige Verfahren noch einsetzbar ist bzw. erlaubte Verdeckungsrate,
- minimale Pixelabdeckung einer Person, bei der das jeweilige Verfahren noch zwischen einzelnen Personen unterscheiden kann,
- erlaubte Blickwinkel der Erfassungstechnik,
- Größe des erfassbaren Bereichs,
- erfasste Daten (z. B. makroskopisch oder mikroskopisch bzw. Trajektorien oder Ein- und Austrittszeiten),
- Generalisierung und Automatisierungsgrad (z. B. benötigt Kalibrierung, Einlernen, manuelle Eingriffe oder halbautomatische Verfahren),
- örtliche Auflösung bzw. Genauigkeit der detektierten Position,
- zeitliche Auflösung der ermittelten Laufwege,
- Einsatzmöglichkeit unter Laborbedingungen (*ex situ*) oder für Feldstudien (*in situ*),
- Art der Erfassungstechnik (z. B. Kamera oder Laserscanner),
- erlaubte Lichtverhältnisse (z. B. im Freien mit von der Sonne ausgesandtem Infrarotlicht oder im Dunkeln),
- erlaubte Umgebungsbedingungen (z. B. Abschattung durch Wände, Rauch, Regen, Nebel, Wolken)
- unbemerkte Erfassung oder Abhängigkeit von der Aktivität der zu erfassenden Personen (z. B. App-Installation, Markierung, Instrumentierung),
- passive oder aktive Marker bzw. Erfassung der Marker (z. B. Codes oder RFID-Chips),
- passive oder aktive Sensoren bzw. Eingriff in die Szene (z. B. Kamera oder Laserscanner),
- statische oder mobile Sensoren (z. B. Überwachungskamera oder mobiler Roboter),

- Reichweite und Auflösung der Sensoren (z. B. Laserscanner oder Kamera),
- Echtzeitfähigkeit (z. B. Robotik oder wissenschaftliche Bewegungsanalyse),
- systembedingte Quote der erkannten Personen (z. B. Bluetooth-Scanner oder GPS-Ortung),
- Einsatzmöglichkeiten (z. B. große Menschenmengen oder detaillierte Bewegungsstudie).

Ich habe mich für eine Gliederung der Verfahren nach der verwendeten Hardware entschieden, da dies eine klare Unterteilung erlaubt und auch die meisten der verwendeten Merkmale abhängig von der verwendeten Technik zur Erfassung sind. Bei den ausführlicher betrachteten Verfahren unter Verwendung von Kameras wird eine weitere Unterteilung nach Merkmalen durchgeführt.

Es gibt eine Vielzahl von Übersichtsartikeln, Sammlungen und Büchern zur Personenerfassung, die die große Menge an Arbeiten in dem Gebiet der Personenerfassung und des Verfolgens (vgl. Anhang B) zu gliedern versuchen. Eine Auswahl dieser Arbeiten sortiert nach ihrem Erscheinungsdatum:

- The Visual Analysis of Human Movement: A Survey [96] (1998):
Übersicht von Verfahren zur Erfassung von Personen mit monokularen Kameras auf Basis von 2D- und 3D-Modellen,
- A survey of computer vision-based human motion capture [97] (2001):
Übersicht von Verfahren zur Erfassung von Personen und ihren Laufwegen in unterschiedlichen Detailstufen,
- Object Tracking: A Survey [98] (2006):
Übersicht von Verfahren zur Erfassung und zum Tracken von Objekten mit einem Schwerpunkt auf der Personenverfolgung,
- A survey of techniques for human detection from video [99] (2006):
Übersicht von Verfahren zur Personendetektion mit monokularen Kameras,
- Crowd analysis: a survey [100] (2008):
Übersicht von Verfahren zur Erfassung mit monokularen Kameras und Analyse von Menschenmengen u. a. durch Modelle der Fußgängerdynamik,
- Survey of Appearance-Based Methods for Object Recognition [101] (2008):
Kurzeinführung in die Funktionsweise generativer Modelle zur Personendetektion,
- Special Issue on Anthropocentric Video Analysis: Tools and Applications [102] (2008):
Sammlung von Artikeln zur Analyse von detaillierten Personeneigenschaften anhand von Videodaten,
- Pedestrian detection and tracking using stereo vision techniques [103] (2008):
Übersicht von Verfahren zur Personendetektion und -verfolgung mit monokularen und Stereokameras und Vorstellung eines eigenen Stereo-Verfahrens,
- Pedestrian Behavior: Models, Data Collection and Applications [104] (2009):
Buch zur Modellierung und Erfassung der Fußgängerdynamik,
- Visual people detection - different models, comparison and discussion [105] (2009):
Übersicht von Verfahren zur Personendetektion mit monokularen Kameras und der Kombination mit Laserscanner-Daten zur Verbesserung der Ergebnisse,

- Monocular Pedestrian Detection: Survey and Experiments [106] (2009):
Übersicht von Verfahren zur Personendetektion mit monokularen Kameras und Untersuchung der Güte von Implementationen dieser Verfahren auf Testdatensätzen,
- A Survey of Human-Sensing: Methods for Detecting Presence, Count, Location, Track, and Identity [33] (2010):
Übersicht von Techniken zur Erfassung von Personen in unterschiedlichen Detailstufen,
- Crowd Analysis Using Computer Vision Techniques [107] (2010):
Übersicht von Verfahren zur Erfassung von Personen in hohen Personendichten mit monokularen Kameras in unterschiedlichen Detailstufen,
- Visual Analysis of Humans – Looking at People [108] (2011):
Buch mit Übersichtsartikeln zur Personendetektion unterschiedlicher Detailstufen und ihrer Anwendung,
- Pedestrian detection: An evaluation of the state of the art [109] (2012):
Vorstellung und Bewertung von aktuellen Implementationen zur Personenerkennung,
- Human detection from images and videos [110] (2012):
Übersicht von Verfahren zur Personendetektion mit monokularen Kameras und Vorstellung eines neuen Verfahrens,
- Special Issue on Behaviours in Video [111] (2013):
Sammlung von Artikeln zur Erfassung und Analyse von Personen in Videodaten,
- Modeling, Simulation and Visual Analysis of Crowds [112] (2013):
Buch mit Übersichtsartikeln zur Erfassung, Analyse und Modellierung von Menschenmengen,
- A Literature Review on Video Analytics of Crowded Scenes [113] (2013):
Übersicht von Verfahren zur Erfassung und Analyse von Menschenmengen,
- Special Issue on Pattern Recognition and Crowd Analysis [114] (2014):
Sammlung von Artikeln eines Workshops zur Erfassung und Analyse von Menschenmengen.

Aufgrund der vielfältigen kommerziellen Anwendungen gibt es eine Reihe von Firmen, die Produkte zur Erfassung von Personen anbieten. Dabei unterscheidet sich der Anspruch und die Qualität der Ergebnisse erheblich voneinander. Erfassen einige Systeme nur sich bewegende Pixel, deren zusammenhängende Größe ggf. eingestellt werden kann, um zwischen Tieren, Personen und Autos zu unterscheiden, bieten andere die Möglichkeit auch Laufwege in Echtzeit zu erfassen und Heatmaps zur Visualisierung von Aufenthaltswahrscheinlichkeiten zu erzeugen. Die beworbenen Erkennungsraten von oftmals fast 100% sind kritisch zu hinterfragen [13] und sind u. a. stark von Personendichte, Lichtverhältnissen, Homogenität der Menschenmenge und der Kalibrierungsgüte abhängig. Die angegebenen Erkennungsraten sind sogar von nicht echtzeitfähigen in diesem Kapitel vorgestellten Verfahren oftmals nicht erzielbar. Eigene Erfahrungen und Vergleiche sind Abschnitt 3.6.7 zu entnehmen. Listen von Firmen sind, gegliedert nach den verwendeten Techniken, in Anhang E aufgeführt. Die meisten Systeme arbeiten mit Bildern von vom Menschen sichtbaren Licht erfassenden monokularen Kameras, da diese kostengünstig sind und die Nutzung von bereits installierten Überwachungskameras ermöglichen, aber auch Laserscanner, Stereokameras, Tiefenkameras oder Wärmebildkameras kommen zum Einsatz.

Die bei Laborversuchen anderer Forschergruppen eingesetzten Methoden zur Erfassung von Laufwegen sind am Ende jedes Abschnittes der einzelnen verwendeten Technik aufgeführt. Eine Übersicht einer Auswahl dieser Versuche ist Tab. 2.1 zu entnehmen. Die meisten Versuche nutzen Farbmarker zur Lokalisierung einer Person, da dies eine einfache und zuverlässige Detektionsmethode darstellt, was ausführlich im Kapitel 3 erläutert wird. In der Tabelle sind nur solche Versuche aufgeführt, bei denen die Laufwege voll automatisch erfasst wurden. Eine manuelle oder halbautomatische Extraktion wurde u. a. in [12, 115–127] angewandt.

2.3 Erfassungstechnik

Zur Erkennung von Personen gibt es eine Vielzahl von Techniken [33], die je nach Anwendungsfall ihre Berechtigung haben. Zum Beispiel für die Schätzung der Personendichte bei Großveranstaltungen im Hellen sind monokulare Kameras eine gute Wahl, einzelne Personen im Dunkeln zu orten ist besser mit aktiven oder andere Lichtspektren nutzenden Sensoren möglich. Aber auch für das eingegrenzte Ziel der präzisen Erfassung von Trajektorien in Personenströmen hoher Dichte sind je nach Randbedingung (u. a. Lichtverhältnisse, Größenunterschiede, maximale Dichte, Abdeckungsbereich, Möglichkeit der Aufhängung und Ausrichtung) unterschiedliche Techniken anwendbar. Im Folgenden soll auf die Methoden näher eingegangen werden, die von uns oder anderen Personen zum Zwecke der exakten Personendetektion verwendet wurden oder in Zukunft werden sollen (siehe Abschnitt 5.2).

2.3.1 Kamera

Mit Kamera ist hier eine monokulare vom Menschen sichtbares Licht erfassende Kamera gemeint, die auf einem 2D-Sensor wie das menschliche Auge ein perspektivisches Abbild der Umgebung erfasst. Dabei kann der Sensor entweder Farb- oder Grauwerte detektieren, wobei die meisten Verfahren ein Grauwertbild als Eingabe erwarten.

Im Gegensatz zu anderen Sensoren (Tast-, Temperatursensor oder Lichtschranke), bietet eine Kamera die höchste Auflösung aller Sensoren und eine große Menge an möglichen Informationen wie z. B. Größe, Farbe, Form, Textur oder Bewegung, die zur Personendetektion herangezogen werden können. Zudem sind Kameras vergleichsweise günstig, durch die breite Verwendung ausgereift und in einer großen Bandbreite an Variationen verfügbar. Durch die große Verbreitung gibt es für die Analyse von Bild- oder Bewegtbilddaten eine Vielzahl von entwickelten Verfahren, die überdies bei der Entwicklung von Erfahrungen beim eigenen menschlichen Sehen profitieren können. Die hohe Informationsmenge und -dichte hat aber zum Nachteil, dass die Komplexität der Algorithmen im Bereich der Bildanalyse hoch und aufgrund der großen Datenmenge oftmals auch rechenintensiv ist. Da viele Anwendungen bei Nutzung von Sensoren einer Echtzeitfähigkeit bedürfen, ist die Auswahl der einsetzbaren Methoden für derartige Anwendungen auf solche beschränkt, die Bildsequenzen mit der für die Anwendung erforderlichen Bildwiederholrate direkt verarbeiten und das Ergebnis in benötigter Zeit bestimmen können. Aufgrund der Komplexität der Eingangsdaten und damit der Verfahren ist zudem mit Fehlern (z. B. zu viel oder zu wenig erkannte Personen) zu rechnen, die bei der Bewertung der Resultate berücksichtigt werden müssen. Die räumliche Auflösung der Bilddaten ist hoch und erlaubt eine genaue Lokalisation von Objekten. Dabei muss die perspektivische Projektion der 3D-Welt auf den 2D-Sensor beachtet werden, die bei Messaufgaben Zusatzinformation wie z. B. den Abstand und die Orientierung zur Kamera bedarf.

Initiator ¹	Fokus ²	Methode ³	Veröffentl. ⁴
S. P. Hoogendoorn, <i>Delft University of Technology</i>	Engstelle, Korridor	Farbmarker	2003, [133–135]
W. Daamen, <i>Delft University of Technology</i>	Zugang, Drehtür etc.	Farbmarker	2008, [136–138]
W. Daamen, <i>Delft University of Technology</i>	Kapazität an Türen	Farbmarker, IR	2010, [139]
X. Liu, <i>University of Science and Technology of China</i>	Oszillation	Farbmarker	2009, [140]
W. Song, <i>University of Science and Technology of China</i>	Engstelle, Korridor	Farbmarker	2010, [141, 142]
M. Moussaid, <i>Universite Paul Sabatier</i>	Selbstorganisation	Farbmarker	2009, [143]
S. Saadat, <i>Ateneo De Manila University</i>	Gehweg	Farbmarker	2010, [144, 145]
S. C. Wong, <i>University of Hong Kong</i>	Kreuzung	Farbmarker	2010, [146]
G. Ren-Yong, <i>Inner Mongolia University</i>	bidirektionaler Verkehr	Farbmarker	2012, [147]
M. Bukáček, <i>Czech Techn. Univ. in Prague</i>	Engstelle	strukt. Marker	2013, [148]
K. Christensen, <i>Utah State University</i>	Menschen mit Behinderung	Codemarker	2013, [149–151]
T. Nescher, <i>ETH Zürich</i>	Labyrinth	Codemarker	2013, [152]
I. Karkin, <i>SITIS Ltd.</i>	Engstelle	Wärmebild	2010, [153]
S. Hostikka, <i>VTT Technical Research Centre of Finland</i>	Gebäuderäumung	RFID	2007, [13, 154]
S. Lemerrier, <i>INRIA Rennes</i>	geschlossener Korridor	Infrarotlicht	2011, [155–157]
D. Bauer, <i>University College London</i>	Laborplattform pamela	Laserscanner	2008, [158, 159]

Tab. 2.1: Laborversuche anderer Forschergruppen mit automatischer Erfassung der Personen, die nicht meine Verfahren zur Extraktion der Laufwege verwenden; ¹Person, die die Versuchsdurchführung initiiert oder die Laufwege extrahiert hat, sowie deren Institution, ²inhaltlicher Schwerpunkt oder Geometrie, ³Verfahren oder Technik zur Detektion der Personen,

⁴Jahr der Erstveröffentlichung und Veröffentlichungen, in denen auf die Erkennung eingegangen wird

Da Kameras eine große Menge an intuitiven Bilddaten erfassen können, ist zur Interpretation des Gesehenen eine manuelle Sichtung der Daten oftmals eine flexible und schnelle Lösung. Erst bei einer Vielzahl von zu erfassenden Objekten (große Menschenmengen), langen Sichtungen (z. B. Überwachungskameras), einer schnellen Verarbeitung (z. B. Produktionsstraßen in der Industrie) oder einer exakten Auswertung (z. B. Auslesen von QR-Codes, Messungen) ist der Mensch überfordert, obwohl das Sehen das die bewussten Sinneseindrücke dominierende Sinnesorgan darstellt [160, 161]. Bei komplexen Teilanwendungen ist eine halbautomatische Auswertung möglich, wie dies z. B. bei der Extraktion von Laufwegen bei unseren ersten Versuchsreihen realisiert oder in [162] für die Personenzählung in großen Menschenmengen aus Luftaufnahmen verwendet wurde. Bei unseren ersten Experimenten wurde die anspruchsvolle Erkennung einer Person manuell und das anschließende Verfolgen über die Bildsequenz automatisch durchgeführt. Auch ist die Bildserie in Abb. 2.3 a noch ein Hilfsmittel zum manuellen Studium der Bewegungsabläufe des Körpers, da der Ablauf aufgrund der Geschwindigkeit vom menschlichen Auge nicht oder nur schwer erfasst werden kann. Bereits 1933 wurden zur systematischen Analyse von Bewegungen im Verkehr von Greenshield Bildserien aufgenommen, die ebenfalls im Anschluss manuell ausgewertet wurden [163] (siehe Abb. 2.4). Er begründete mit seiner Arbeit die Beschreibung des Verkehrs mit Hilfe von Fluss- oder Dichte-Geschwindigkeits-Diagrammen, den sogenannten Fundamentaldiagrammen [164] (vgl. Abb. 1.4).

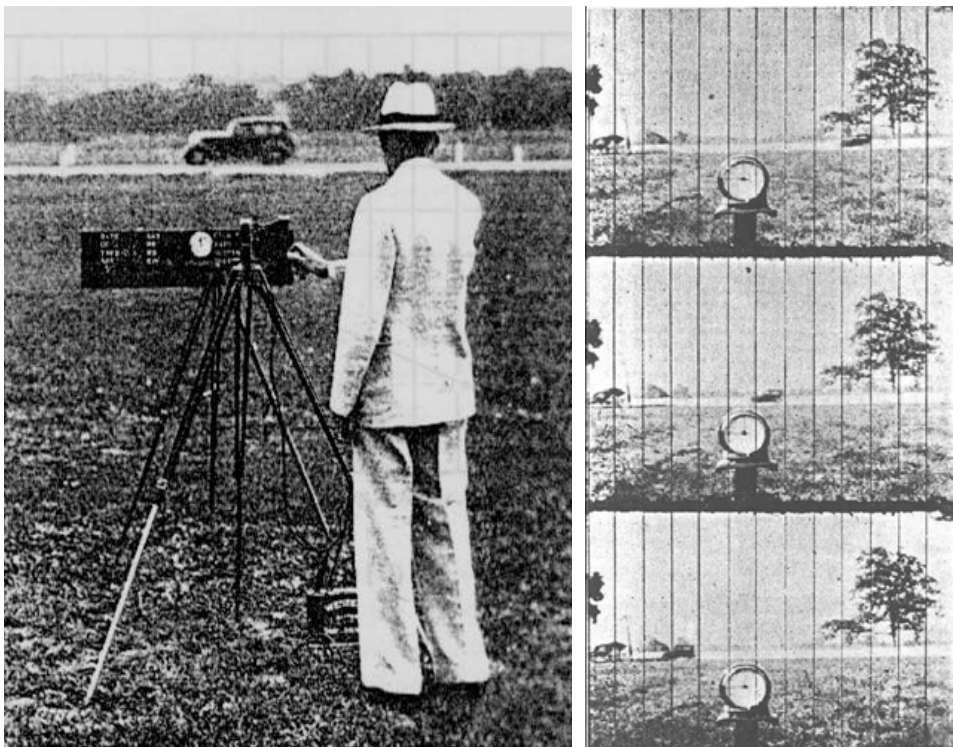


Abb. 2.4: Bildserie mit zeitlich konstantem Einzelbildabstand aufgenommen von im linken Bild stehenden Greenshield mit Hilfe einer batteriebetriebenen Filmkamera; eine eingblendete Uhr diente der Zeitmessung und überlagerte vertikale Linien der Entfernungsmessung von zur Bildebene parallelen Bewegungen vorbeifahrender Autos; Bilder aus [163]

2.3.1.1 Hintergrundsubtraktion

Viele Verfahren versuchen im ersten Schritt Vorder- und Hintergrund zu segmentieren. Dies stellt für Bildsequenzen einer sich bewegenden Kamera eine noch größere Herausforderung als für Sequenzen einer statischen Kamera dar. Aber auch bei Kameras mit einer festen Perspektive gibt es Herausforderung, z. B. in Fällen, in denen

- die Szene meist mit Vordergrundobjekten belegt ist,
- die Lichtverhältnisse sich ändern,
- sich Objekte im Hintergrund bewegen, wie z. B. Wolken oder Bäume im Wind oder
- es Schattenwürfe gibt, die nicht den Schatten werfenden Objekten zugeordnet werden sollen.

Zwei Klassen von Verfahren können hierbei unterschieden werden. Die verbreitetste ist die Erzeugung eines Hintergrund-Modells, das Farbe, Helligkeitsgradienten, Kanten, Textur oder andere Merkmale (*features*) über die Zeit sammelt und aktualisiert. Die zweite Gruppe von Methoden geht von einer statischen Kamera aus und klassifiziert diejenigen Pixel als Hintergrund, die sich nicht bewegen.

Die erste Klasse deklariert dann einen Bereich des Bildes als Vordergrund, wenn das Bild in diesem Bereich sich von den Hintergrund-Merkmalen unterscheidet. Die Merkmale können sich über die Zeit z. B. durch Änderung der Lichtverhältnisse verschieben, so dass bei längeren Beobachtungen das Modell angepasst werden muss. Bei kurzen Sequenzen wie bei den von mir auszuwertenden Versuchen reicht es hingegen, dass pro Pixel kurz vor dem Beginn des Versuchs der durchschnittliche Farbwert der leeren Szene ermittelt und damit der gesamte Durchlauf verglichen wird. Bei einer Aktualisierung des Hintergrund-Modells muss beachtet werden, wie der Vordergrund sich über die Zeit verändert. Bedeckt der Vordergrund Bereiche des Hintergrundes über eine lange Zeit oder bleiben Objekte stehen, könnte bei einer schnellen Anpassung des Hintergrundes der Vordergrund dem Hintergrund zugeschrieben werden. Eine häufige Art der Modellierung der Varianz des Hintergrundes nutzt zusammengesetzte Gauß-Verteilungen (*mixture of Gaussians*) [165, 166]. Verfahren, die die Farbe als Merkmal nutzen sind in [167, 168] zu finden, [169] nutzt Gradienten und [170, 171] zieht die Lage der Kanten in die Unterscheidung zwischen Vorder- und Hintergrund ein.

Die zweite, seltener verwandte Klasse zur Unterscheidung zwischen Vorder- und Hintergrund nutzt den optischen Fluss [23, 172]. Hierdurch sind jedoch nur Objekte im Vordergrund identifizierbar, die sich bewegen und dies nicht entlang der optischen Achse tun. Dafür sind sich unabhängig voneinander bewegende Objekte innerhalb des Vordergrundes segmentierbar [173], können aber dadurch z. B. auch Arme oder Beine eines Körpers fälschlich als getrennte Objekte erkennen. Der optische Fluss ist unabhängig von langfristigen Änderungen in der Beleuchtung, jedoch kann die Bewegung eines Schattens ebenfalls als Bewegung detektiert werden. Zudem ist zur Bestimmung des Flusses eine texturierte Oberfläche nötig.

2.3.1.2 Klecks-Analyse

Nachdem der Vordergrund bestimmt wurde, können diese Bereiche analysiert werden. Die direkte pixelbasierte Klecks-Analyse (*blob analysis*) verwendet direkt die zusammenhängenden Gebiete des Vordergrundes. In den Anfängen der automatischen Verfahren der computergestützten Bildanalysen zur Personendetektion wurde davon ausgegangen, dass jeder Klecks einer Person entspricht [174],

später wurden Bedingungen an den Flächeninhalt oder die Form eines Kleckses gestellt, um Personen z. B. von Autos unterscheiden zu können. Die Form, Größe und Anzahl kann auch während des Trackings analysiert werden, um sich trennende Personen als solche separat zu erfassen [175, 176].

Eine echtzeitfähige verlässliche Personendetektion mittels Klecks-Analyse ist am ehesten mit überkopf angebrachten senkrecht zum Boden gerichteten Ansichten von Kameras mit einer großen Brennweite möglich [177, 178]. Dies ist z. B. für Zählkameras an Ein- und Ausgängen ein realistisches Szenario (vgl. Anhang E), wobei eine große Brennweite bei geringer Deckenhöhe oft nicht zu erfüllen ist.

Wichtig ist bei der Klecks-Analyse eine gute Hintergrundsubtraktion, die Schatten als solche erkennt und diese nicht dem Vordergrund zuschreibt, da der Schatten eines Menschen der Form eines Menschen entspricht und damit als solcher gezählt würde.

Pixelbasierte Verfahren wie die Klecks-Analyse arbeiten in kontrollierter Umgebung gut, erlauben eine detaillierte Segmentierung der Personen und sind schnell, wohingegen die im Folgenden vorgestellten objektbasierten Verfahren wie der Vorlagen-Abgleich oder die Nutzung eines Form-Modells ihre Stärken bei komplexen Szenen mit hohen Personendichten, Verdeckungen und sich ändernden Lichtverhältnissen haben [108].

2.3.1.3 Vorlagen-Abgleich

Beim Vorlagen-Abgleich (*template matching*) werden eine oder mehrere Vorlagen einer ganzen Person oder von Teilen einer Person mit dem Bildinhalt verglichen. Dabei muss der Grad der Übereinstimmung mit der Vorlage für das ganze Bild in verschiedenen Größen und Orientierungen bewertet werden. Die Vorlage sollte so generell wie möglich sein, um unterschiedliche Posen abzudecken, aber so speziell, dass nicht zu viele falsch positive Detektionen vorkommen. Dabei sollte die Vorlage entsprechend der Ausrichtung der Kamera gewählt sein. Sie unterscheidet sich damit z. B. stark für überkopf angebrachte Kameras und Kameras mit einer seitlichen Sicht. Als Ergebnis erhält man einen Bereich, in dem die Vorlage entsprechend der Bewertung passt und nicht wie im vorangegangenen Abschnitt eine Kontur, die die Person umschreibt. Der Test auf Übereinstimmung kann sehr rechenintensiv sein, da ggf. mehrere Posen mit einem sich verschiebenden Fenster (*sliding window*) über das ganze Bild auf unterschiedlichen Skalen und mit variierenden Orientierungen verglichen werden müssen. Zur Beschleunigung kann das Bild zuerst in einer geringeren Auflösung abgetastet werden und nur für Bereiche mit möglichen Kandidaten die Auflösung erhöht werden, oder die Vorlagen bilden eine Hierarchie, bei denen tieferliegende Hierarchiestufen erst untersucht werden, wenn die darüber liegende der Vorlage ungefähr entspricht (siehe Abb. 2.5 a) [179, 183, 184].

Statt viele feste Vorlagen zu nutzen, wurden Verfahren entwickelt, die ein einziges statistisches Modell (*statistical shape model*) nutzen, das mögliche Variationen anhand von Orientierungspunkten (*landmarks*) enthält (*point distribution model*, PDM) [185, 186]. Der anhand von Trainingsdaten ermittelte Raum der Verteilungen dieser Punkte wird oft mittels Hauptkomponentenanalyse bzw. Singulärwertzerlegung (*principal components analysis*, PCA) für den eigentlichen Test vom Umfang reduziert. Zur effizienten Lokalisierung einer Person und optimalen Anpassung der parametrisierten Vorlage kann bereits nach einem ersten ungefähren Treffer das Modell über die Orientierungspunkte auf die beste Passform gebracht werden (*active shape model*, ASM) [187–189].

Statt den gesamten Körper oder große Teile wie den häufiger zu sehenden Kopf-Schulter-Bereich im Bild zu suchen, können auch nur kleine Körperteile (*low level features*) gesucht und im Ganzen bewertet werden. Alle Verfahren, die Klassifikatoren (*classifier*) nutzen, um lokal zu entscheiden, ob

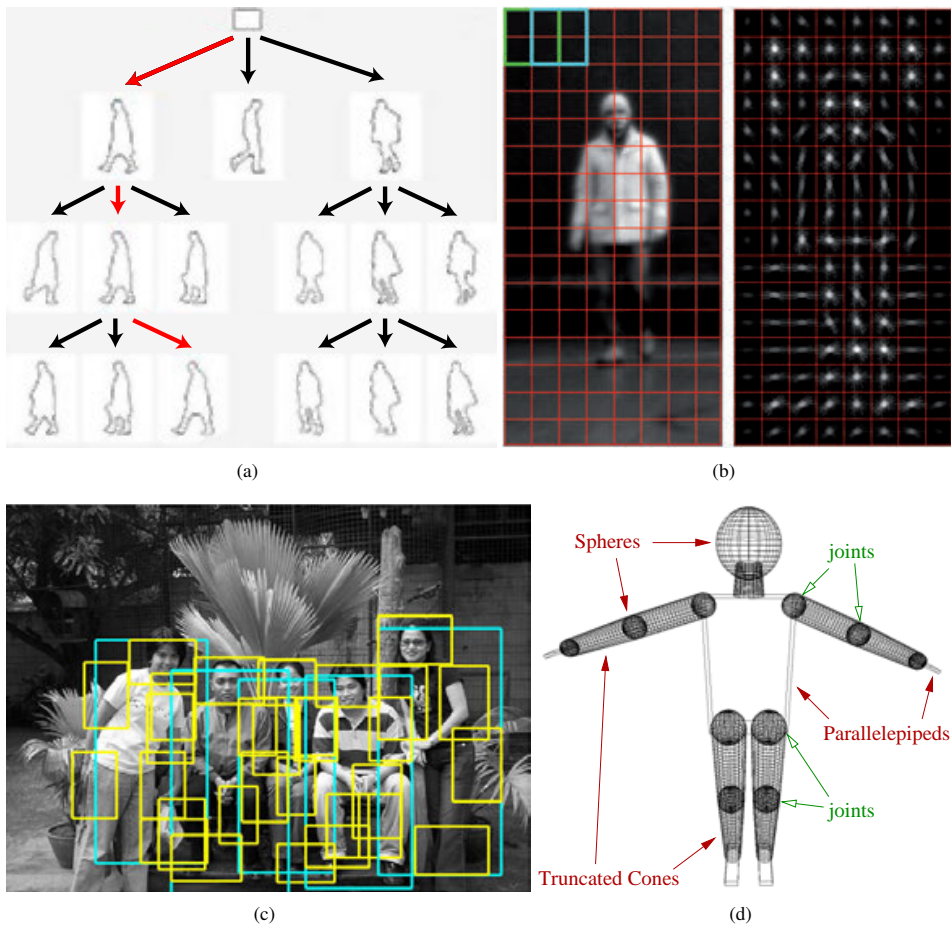


Abb. 2.5: Veranschaulichungen von Verfahren und Ergebnissen zur Personendetektion mittels monokularer Kameras; (a) Hierarchiestufen für den Vorlagen-Abgleich aus [179]; (b) Regelmäßiges Gitter mit Histogrammen orientierter Gradienten (*histogram of oriented gradients*, HOG) des linken Bildes; (c) detektierte Personen mittels verformbare Teile basierendem Modell (*deformable part-based model*, DPM) aus [180]; (d) 3D-Modell mit 20 Freiheitsgraden, das in [181] als 3D-Form einer Person benutzt wurde; Bilder aus (a) [179], (b) [182], (c) [180], (d) [181]

es sich um eine Person handelt oder nicht, können Verfahren des maschinellen Lernens (*machine learning*) nutzen. Bekannte Verfahren hierfür sind die Stützvektormethode (*support vector machine*, SVM) [190], das angepasste Verstärken (*adaptive boosting*, AdaBoost) [193] oder das überwachte Lernen (*supervised learning*) mittels unkorrelierter Entscheidungsbäume (*random forest classifier*) [194] oder künstlicher neuronaler Netze (*artificial neural network*, ANN) [49, 195]. Zur Beschleunigung der großen Anzahl an Tests werden auch hier oftmals hierarchische Ansätze, sogenannte Kaskaden, eingesetzt, so dass weitergehende Analysen nur an Orten durchgeführt werden, an denen durch vorangegangene undetailliertere Tests der Ort einer möglichen Person angezeigt wurde [196–199]. Diese Art der Personenerkennung ist für monokulare Kameras die verbreitetste Methode zur Detektion

einer Person, u. a. da sie keine Hintergrundsubtraktion voraussetzt und mit sich ändernden Beleuchtungssituationen umgehen kann. Dabei sind Verfahren, die ein Histogramm orientierter Gradienten (*histogram of oriented gradients*, HOG) [200] als Merkmal zur Beschreibung einer Person einsetzen, aktuell die erfolgreichsten [201]. Die HOG-Beschreibung erweitert die Beschreibung gemäß der skalenunabhängigen Merkmalstransformation (*scale-invariant feature transform*, SIFT) [202] dahingehend, dass das zu untersuchende Fenster in ein regelmäßiges Gitter mit Zellen der Größe $4\text{ px} \times 4\text{ px}$ oder $8\text{ px} \times 8\text{ px}$ unterteilt wird und ein Histogramm ungerichteter Gradienten ($\in [0, 180^\circ]$) jedes Pixels für sich überlappende 2×2 Zellenblöcke erstellt wird (siehe Abb. 2.5 b). Hierarchische Erweiterungen, die erst in tieferen Stufen das zu untersuchende Fenster detailliert auf markante Körperteile testen, sind z. B. das auf verformbaren Teilen basierende Modell (*deformable part-based model*, DPM) [180] (siehe Abb. 2.5 c) oder das implizite Formmodell (*implicit shape model*, ISM) [204, 205]. [206] ähnelt dem in [205] vorgestellten impliziten Formmodell, nutzt bei der Erkennung einer Person aber statt der Form die mittels [207] extrahierte Bewegung als Merkmal, so dass stehende Personen z. B. in dichten Menschenmengen nicht erkannt werden können.

2.3.1.4 3D-Form-Modell

Bisher wurde direkt in der Bildebene nach Formen oder Teilen gesucht, die ebenfalls in der Bildebene oft durch Lernstrategien gewonnen wurden. Bei Verfahren, die ein 3D-Form-Modell (*3D shape model*) (siehe Abb. 2.5 d) einsetzen, wird ein einziges dreidimensionales Personenmodell parametrisiert und werden Ansichten dieses virtuellen Modells gesucht [195, 208]. Dies hat den Vorteil, dass beim Finden einer Person nicht nur der Bereich identifiziert wird, in dem sich eine Person aufhält, sondern auch eine Aussage über die Lage der Körperteile gemacht werden kann.

Für ein präzises kinematisches Modell des menschlichen Körpers werden mindestens 20 Freiheitsgrade (*degree of freedom*, DOF) benötigt [209]. Mit den zusätzlichen Freiheitsgraden der Blickrichtung der Kamera ist der Suchraum sehr groß und damit die Rechenzeit hoch, so dass viele Verfahren Annahmen treffen, die die Anzahl der Freiheitsgrade herabsenkt, u. a. indem die initiale Pose der Person als bekannt vorausgesetzt wird [181]. Die Suche nach passenden Parametern kann mit Hilfe von Monte-Carlo-Methoden (*Monte Carlo method*) erleichtert werden, z. B. nutzt [209] Markow-Ketten (*Markov chain*) zur Steuerung der Monte-Carlo-Suche.

Verfahren basierend auf 3D-Form-Modellen ähneln statistischen zweidimensionalen Modellen zur Beschreibung der Körperform in einem Bild, nur dass statt der Verteilung der Orientierungspunkte der Vorlage die beschriebenen Freiheitsgrade als Parameter zur Beschreibung der möglichen Körperformen genutzt werden.

Für alle bisher vorgestellten Techniken gilt, dass eine Verbesserung durch die gleichzeitige Verwendung von mehreren Verfahren (*multi-cue approach*) [210–215] oder die Nutzung von zeitlichen Informationen [216–218], wie die Frequenz oder Bewegungsmuster von sich wiederholenden Bewegungen z. B. der Beine [206, 219–223], die aus Bildsequenzen gewonnen werden können, erzielt werden kann. Dabei sinkt für gewöhnlich die Falsch-Ermittlungs-Rate, aber es steigt im Gegenzug die Falsch-Negativ-Rate, da nicht alle gesuchten Merkmale für jede Person zutreffen, z. B. wird die Ganganalyse bei stehenden Personen oder die Gesichtserkennung bei abgewendeten Personen fehlschlagen. Vorwissen über die Szene kann ebenfalls eine verlässlichere Detektion ermöglichen. Im folgenden Abschnitt wird motiviert, warum ich die Detektion mit Vorwissen durchführe und in welchen anderen Experimenten dies ebenfalls genutzt wurde.

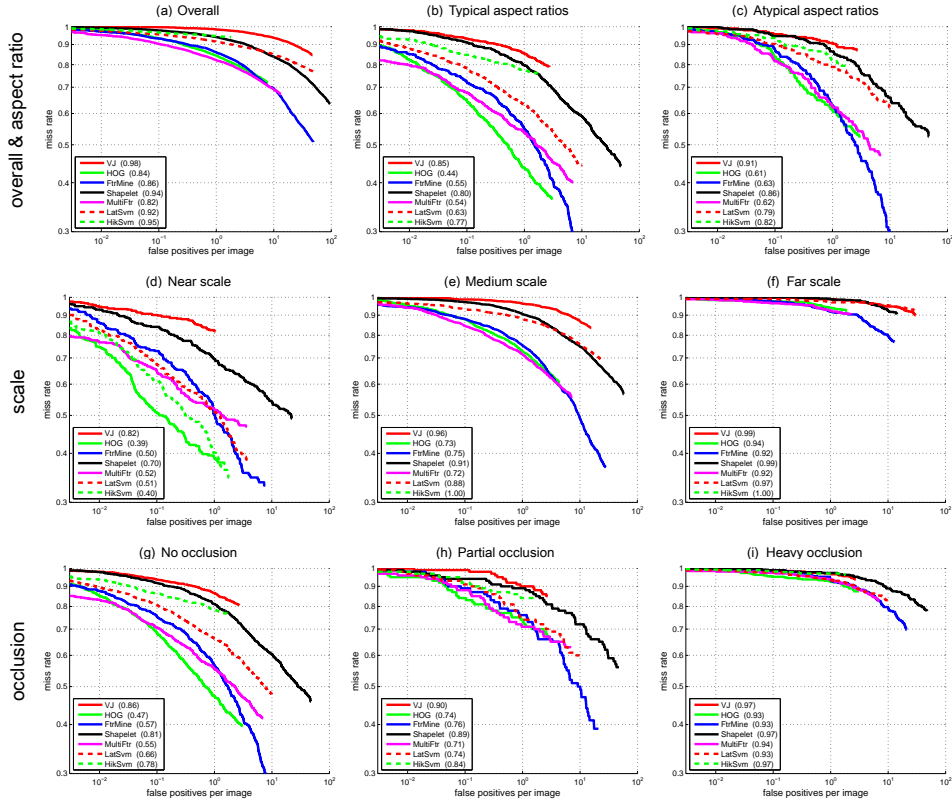


Abb. 2.6: Falsch-Negativ-Rate (siehe Abschnitt 3.6) im Verhältnis zu falsch positiven Detektionen pro Bild für aktuelle Implementierungen verbreiteter Verfahren zur Personendetektion und für unterschiedliche Szenarien: (b) typisches oder (c) atypisches Seitenverhältnis der abgebildeten Personen, (d-f) Pixelabdeckung pro Person, (g-i) Grad der Verdeckung; in der Legende ist in Klammern die Falsch-Negativ-Rate für eine falsch positive Detektion pro Bild angegeben; weitere Verfahren und Untersuchungen an verschiedenen Testdatensätzen in [109]; Diagramme aus [224]

2.3.1.5 Detektion mit Vorwissen

Abb. 2.6 zeigt Diagramme aus [224], die einen qualitativen Eindruck der Erkennungsraten von aktuellen Verfahren geben sollen, ohne im Detail darauf einzugehen, wie die verwendeten Datensätze für die einzelnen Erkennungsläufe genau zusammengesetzt sind. Für genaue Angaben hierzu sei auf [224] verwiesen. Je tiefer die Kurve liegt, desto besser ist das Ergebnis des Verfahrens. Die Falsch-Negativ-Rate liegt für alle Fälle über 30% und erreicht diese Rate nur mit dem FtrMine-Verfahren für einige Szenarien bei ca. 10 falsch positiven Detektionen pro Bild. Ein verbindendes Tracking wird die Falsch-Negativ-Rate verringern, jedoch ist die fehlende Erkennung systematisch und übersieht u. a. Personen, die nah beieinander stehen, klein sind, z. B. Kinder, oder sich nicht stark vom Hintergrund abheben. Aber auch für Szenarien, auf die die Verfahren hin optimiert wurden, liegt die Falsch-Negativ-Rate noch über 10%, so z. B. bei [65, 70, 70, 144, 204, 225–227]. Auch Tab. 2.2 basierend auf einer Tabelle aus [109] dient dazu, einen Überblick über verwendete Techniken und eine Vorstellung von

	Features						Learning			Detection Details					Implementation				
	gradient hist.	gradients	grayscale	color	texture	self-similarity	motion	classifier	feature learn.	part based	non-maximum suppression	model height (in pixels)	scales per octave	frames per second (fps)	log-average miss rate	training data	original code	full image evaluation	publication
VJ								AdaBoost			MS	96	~14	.447	95%	INRIA			'04
SHAPELET		✓						AdaBoost	✓		MS	96	~14	.051	91%	INRIA			'07
POSEINV		✓						AdaBoost			MS	96	~18	.474	86%	INRIA	✓		'08
LATSVM-V1		✓						latent SVM		✓	PM	80	10	.392	80%	PASCAL		✓	'08
FTRMINE		✓			✓			AdaBoost	✓		PM	100	4	.080	74%	INRIA	✓		'07
HIKSVM		✓						HIK SVM			MS	96	8	.185	73%	INRIA	✓		'08
HOG		✓						linear SVM			MS	96	~14	.239	68%	INRIA	✓		'05
MULTIFTR		✓			✓			AdaBoost			MS	96	~14	.072	68%	INRIA	✓	✓	'08
HOGLEP		✓				✓		linear SVM			MS	96	14	.062	68%	INRIA	✓	✓	'09
LATSVM-V2		✓						latent SVM		✓	PM	96	10	.629	63%	INRIA	✓	✓	'09
PLS		✓			✓	✓		PLS+QDA	✓		PM*	96	~10	.018	62%	INRIA	✓	✓	'09
MULTIFTR+CSS		✓				✓		linear SVM			MS	96	~14	.027	61%	TUD-MP	✓	✓	'10
FEATSYNTH		✓			✓			linear SVM	✓	✓	-	96	-	-	60%	INRIA	✓	✓	'10
FPDW		✓						AdaBoost			PM*	100	10	6.492	57%	INRIA	✓	✓	'10
CHN FTRS		✓			✓			AdaBoost			PM*	100	10	1.183	56%	INRIA	✓	✓	'09
MULTIFTR+MOTION		✓				✓	✓	linear SVM			MS	96	~14	.020	51%	TUD-MP	✓	✓	'10

Tab. 2.2: Vergleich von aktuellen Verfahren zur Personendetektion u. a. mit Angabe verwendeter Merkmale (*features*), Lernstrategien (*learning*), der Verarbeitungsgeschwindigkeit ersichtlich an der Bildwiederholrate (*frames per second*) für 640px × 480px große Bilder und einer durchschnittlichen über das geometrische Mittel bestimmte Falsch-Negativ-Rate (siehe Abschnitt 3.6) (*log-average miss rate*) von im Bild klar zu sehender Personen mehrerer Datensätze; die Verfahren konnten von den Autoren auf die Datensätze trainiert und parametrisiert werden; weitergehende Beschreibungen können [109] entnommen werden; basiert auf Tabelle aus [109]

der Performance hinsichtlich Geschwindigkeit und Fehlerrate aktueller Verfahren zur Detektion von Personen zu bekommen.

Da diese Arbeit ein Verfahren benötigt, das auch bei hohen Personendichten funktioniert, sollen hier einige Arbeiten, die speziell die Analyse von Menschenmengen hoher Dichte beschreiben, aufgeführt werden: [38, 60, 65, 68–70, 72, 77, 100, 107, 112–114, 162, 189, 204, 209, 210, 226–242]. Dabei variiert die behandelte Dichte (zwischen lokal $1/\text{m}^2$ und global im gesamten Bild $4/\text{m}^2$) und die Pixelabdeckung je Person (Überwachungskamera im Raum oder Luftaufnahme). Speziell für überkopf ausgerichtete Kameras wie bei unseren Versuchen werden Verfahren in [167, 170, 177, 178, 243–246] beschrieben. Die Ergebnisse sind aufgrund der hohen Dichte oft noch ungenauer als die bereits vorgestellten Methoden bei geringerer Dichte (vgl. Abb. 2.6). Welches Verfahren zu welchem Szenario am besten passt, ist sehr stark vom Blickwinkel und der vorherrschenden Personendichte abhängig. Ein für alle Situationen optimales Verfahren existiert nicht.

Die hohe Fehlerrate der existierenden Verfahren war der Anlass für die eigene Entwicklung, bei der für die Detektion mit monokularen Kameras die Personen mit Markern versehen wurden und für die markerlose Detektion Stereokameras eingesetzt wurden, wie sie im folgenden Abschnitt vorgestellt werden. Für unseren Zweck, der Analyse der Laufwege für quantitativ verlässliche Daten in der Fußgängerdynamik, sind nahezu fehlerfreie Trajektorien jeder Person nötig (siehe Abschnitt 3.6.9), so dass bei den hohen Fehlerraten der existierenden markerlosen Verfahren für monokulare Kameras sehr viel nachträgliche manuelle Arbeit nötig gewesen wäre. Allgemein werden Verfahren, die Vorwissen zur Detektion von Objekten nutzen, unter dem Begriff „Erkennung mit Vorabinformationen“ (*detection with prior information*) geführt, wobei dazu z. B. bereits das Wissen über die Hauptbewegungsrichtung gehört.

Auch andere Forschergruppen haben in Versuchen für die automatische Extraktion von Laufwegen aus Bildsequenzen die Probanden oder in einem Fall die Umgebung mit Markern versehen (siehe Abb. 2.7).

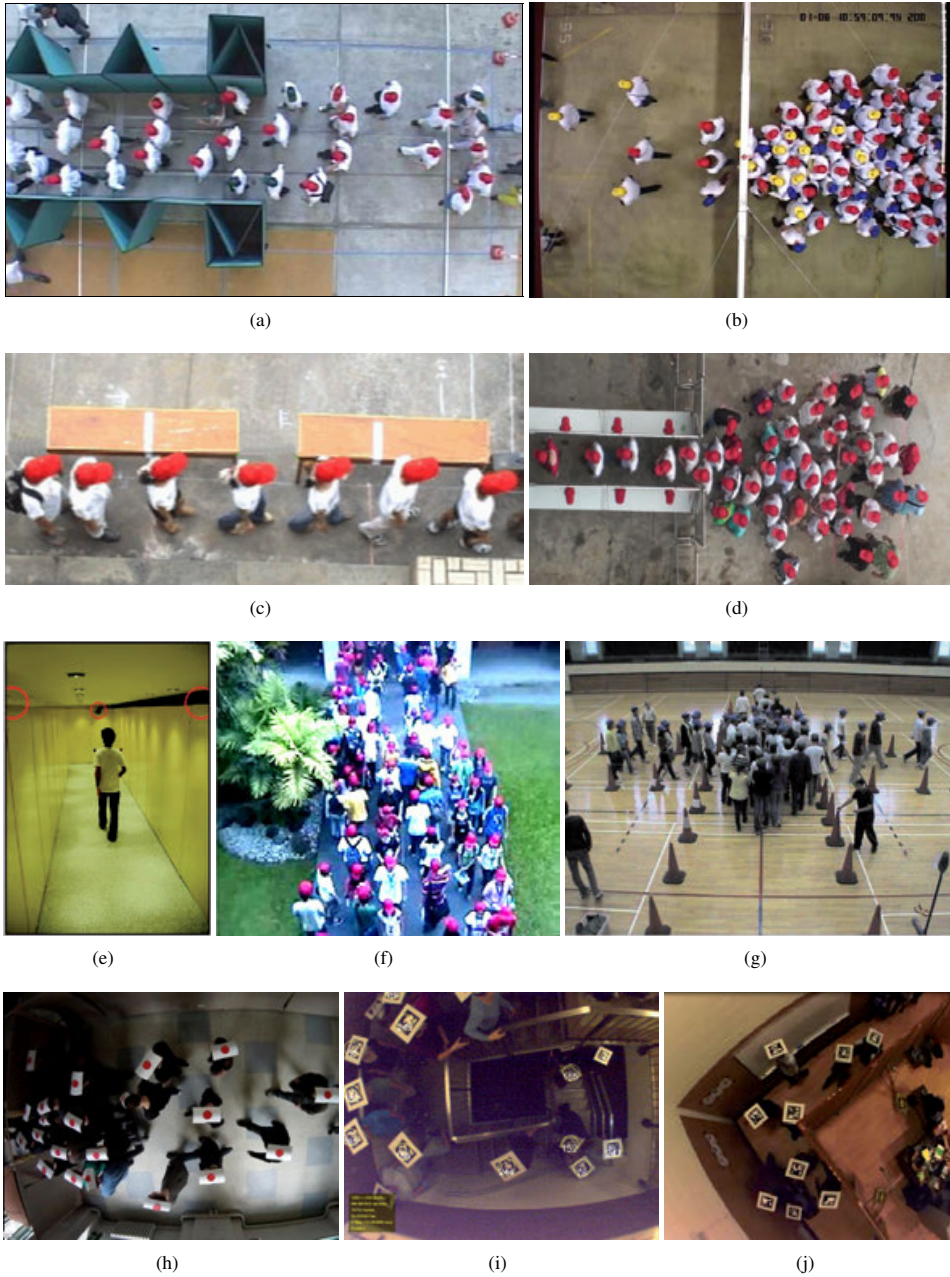


Abb. 2.7: Experimente anderer Gruppen (siehe Tab. 2.1), bei denen monokulare Kameras zur Erfassung der Laufwege genutzt wurden; Bilder basieren auf Abbildungen aus (a) [133], (b) [139], (c) [140], (d) [141], (e) [143], (f) [144], (g) [146], (h) [148] und (i, j) [150], denen auch weitere Informationen zum Versuch entnommen werden können

Bei der Auswertung der Versuche aus [133] wurden farbige (rot, grün) Kopfbedeckungen zur Lokalisation von Personen herangezogen (siehe Abb. 2.7 a). Die Detektion wurde u. a. aufgrund des Schattenwurfs der Aufbauten in einem gleichmäßig gut ausgeleuchteten Bereich vor der Engstelle durchgeführt. Ein Pixel wurde einer Kappe bestimmter Farbe zugeordnet, wenn der Farbwert dieses Pixels im RGB-Farbraum in vorbestimmten Intervallen der R-, G- und B-Kanäle (in der Arbeit als *fuzzy sets* bezeichnet) lag. Die Zusammengehörigkeit von farbigen Pixeln wurde darüber hergestellt, dass alle Pixel zu einer Gruppe (*cluster*) bzw. einem Hut gehörten, wenn ihr euklidischer Abstand zueinander einen Schwellwert unterschritten hat. Die Abstandsbestimmung bedarf einer längeren Berechnung als die von mir für die Zusammenführung der Pixel eines Farbmarkers verwendeten morphologischen Operationen (siehe Abschnitt 3.2.2). Zudem ist die Farbwahl durch einen quaderförmigen Bereich im RGB-Farbraum nicht optimal zur Bestimmung von ähnlichen Pixelfarben. Für Aufnahmen der realen Welt eignet sich der von mir für den Farbmarker verwendete HSV-Farbraum besser [247] (vgl. Abb. 3.14).

Die Bestimmung der Position jeder Person wurde in [139] bei normaler Beleuchtung wie in der im vorherigen Absatz beschriebenen Arbeit [133] durchgeführt (siehe Abb. 2.7 b). Für Experimente mit abgedunkeltem Licht wurden die Kappen mit Infrarot-LEDs versehen, die mit Hilfe einer Infrarotkamera detektiert wurden. Die gleiche Autorin hat in [137, 138] ähnliche Verfahren verwandt. Bei Experimenten zur individuellen Interaktion wurden zudem farbige Schultern verfolgt.

[140] nutzt ebenfalls die Farbe (rot) der Kopfbedeckung zur Lokalisierung der Position einer Person (siehe Abb. 2.7 c) und zählt die Pixel zu einem Hut, deren Farbe im RGB-Farbraum einen gewissen euklidischen Abstand zur Farbe Rot nicht überschreiten. Statt eines Quaders werden nun somit Farben innerhalb einer Kugel im Farbraum dem Marker zugewiesen.

In [141] und [142] wird nicht näher auf die Erkennung der roten Hüte eingegangen, die die Probanden zur Detektion tragen mussten (siehe Abb. 2.7 d), sondern auf [140] verwiesen. Stattdessen wird ausführlich die Art des Trackings beschrieben, das mittels eines Mean-Shift-Trackers [248, 249] durchgeführt wurde.

Die bisher vorgestellten Aufnahmen wurden überkopf mit nahezu senkrechtem Blickwinkel zur Bewegungsebene gemacht, um die Verdeckung der Personen untereinander so gering wie möglich zu halten. Da die Kamera nicht hoch genug aufgehängt werden konnte, um den zu erfassenden Bereich einsehen zu können, wurden die Aufnahmen in den folgenden Versuchsreihen schräg durchgeführt, was zu Verdeckungen und ungenauerer Positionsbestimmung wie in [144, 146] oder aufwändigerem Versuchsaufbau wie in [143] führte (siehe Abb. 2.7 e). Die Detektionsmethode von [143] mittels eines Mehrkamerasystems wird in Abschnitt 2.3.3 vorgestellt.

In [144] werden Verfahren mit und ohne Marker beschrieben. Für die verlässliche Detektion auch dichter Menschenmengen wird ein Verfahren vorgestellt, das Personen aufgrund ihrer roten Kopfbedeckung lokalisiert. Hierfür wurden die Probanden gebeten, auf rote Kleidung zu verzichten (siehe Abb. 2.7 f). Zu Beginn wurde eine Hintergrundsubtraktion vorgenommen. Zur Bestimmung der rötlichen Pixel wurde danach vom Rot-Kanal der Grün-Kanal abgezogen und alle Pixel einem Hut zugeordnet, bei dem dieses Resultat einen Schwellwert überstieg. Anschließend wurden zur Beseitigung von Störungen wie bei der Behandlung meiner Farbmarker (siehe Abb. 3.15) noch morphologische Operationen wie Dilatation, Öffnen und Schließen durchgeführt.

In der Arbeit [146] wurde ebenfalls die Farbe der Hüte (blau, grün) genutzt, um initiale Positionen im Bild zu bestimmen, in der sich die Köpfe der Personen befinden können (siehe Abb. 2.7 g). Hierfür wurden mit der in [250] beschriebenen Methode uniforme Segmente gleicher Farbe ermittelt. Der Schwerpunkt der entsprechend farbigen Segmente diente der Positionsbestimmung unter Berücksich-

tigung des Blickwinkels. Aufgrund des schrägen Blickwinkels, der für entfernte Personen nur wenig zum farbigen Hut gehörende Pixel zur Erkennung ließ und des Weiteren daraus Verdeckungen resultierten, wurden die Daten im Anschluss manuell aufwändig korrigiert. Die Nutzung des Schwerpunktes der Segmente als Position der Personen dürfte dabei wegen der seitlichen Sicht der Kopfbedeckungen unter Annahme einer Durchschnittspersonengröße zu nicht dokumentierten Fehlern geführt haben. In [147] wurde dieselbe Technik wie bei [146] genutzt.

Die einzigen Versuche, bei denen unseren Markern mit Strukturelementen ähnelnde Kopfbedeckungen eingesetzt wurden, sind in [148] beschrieben (siehe Abb. 2.7 h). Zur Entzerrung des Bildes der eingesetzten Fischaugenkamera wurde eine Sinustransformation gemäß [251] durchgeführt. Die Erkennung des roten Punktes auf einer weißen Pappe geschah über die Farbe, erlaubte somit entgegen der anderen Farbmarker eine stabilere Positionsbestimmung, da meist der ganze Punkt im Bild zu sehen war.

Die in [149–151] vorgestellten Versuche nutzten Marker und Extraktionsverfahren der ARToolKit-Plus-Bibliothek [252–254], die für Anwendungen aus dem Bereich der erweiterten Realität (*augmented reality*, AR) entwickelt wurde, und besitzen durch die individuellen Marker Laufwege, die bestimmten Personen zugeordnet werden können, wie es ebenfalls mit unserem Codemarker (Nr. 5) (siehe Abb. 3.9 5) geplant ist. Die auf dem Kopf getragenen Marker (siehe Abb. 2.7 i und j) besaßen eine Größe von 30 cm × 30 cm, um diese bei gewählter Optik und Kameraentfernung auslesen zu können. Die ARToolKitPlus-Bibliothek erlaubt bei exakter Kalibrierung der Kamera die Bestimmung der räumlichen Position der Marker und ihrer Orientierung, wobei die letztere Information nicht ver-

wandt wurde. Die Bestimmung der 3D-Position ermöglichte Versuche an Treppen (siehe Abb. 2.7 i). Für eine ausreichende Kamerakalibrierung wurden Verfahren aus [255–257] verwandt.

In [152] wurde der Laufweg und die Kopfausrichtung einer einzelnen Person mit Hilfe der Kamera eines Tracking-Systems der Firma Intersense [258] erfasst, die an einem am Kopf fixierten Bildschirm (*head mounted display*, HMD) angebracht war. Das System kann neben der Position auch die Ausrichtung registrieren, so dass damit zusammen mit dem erfassten Ort auch die Kopfausrichtung zur Verfügung stand. Zur Lokalisierung werden in der Umgebung angebrachte Marker und neben der Kamera eine Verarbeitungseinheit benötigt, die im Versuch in Form eines Laptops mitgeführt wurde (siehe Abb. 2.8).



Abb. 2.8: Person, die durch ein T-förmiges Labyrinth laufen musste; das rote Rechteck diente zur anfänglichen Kalibrierung; die Verarbeitungseinheit für das am kopffixierten Bildschirm befestigte Tracking-System wurde vom Probanden als Rucksack mitgeführt; Bild aus [152]

2.3.2 Stereokamera

Stereokameras besitzen zwei synchron arbeitende Kameras, um mittels der Differenz der Pixelposition gleicher Objekte bzw. der Parallaxe der Betrachtungswinkel über Triangulation die Entfernung zu diesem Objekt bestimmen zu können (siehe Abschnitt 3.4.2). Werden beide Bilder für eine spätere natürliche Stereopsis, dem räumlichen Sehen, erfasst, so wird als Abstand beider Kameras etwa

der Augenabstand von 6–7 cm gewählt. Bei der automatischen Auswertung der Stereodaten ist dieser Abstand, die sogenannte Basislänge, entsprechend den Anforderungen zu wählen. Je größer die Basislänge, desto genauer kann der Abstand bestimmt werden (vgl. Gleichung 3.26). Eine große Basislänge führt jedoch dazu, dass Bereiche nicht von beiden Perspektiven aus gesehen werden oder zumindest anders erscheinen (u. a. spiegelnde Reflexionen). Für die Personendetektion bedeutet dies, dass bei Verfahren, die möglichst genau für einige in beiden Perspektiven sichtbare Objekte die Position bestimmen müssen, die Basislänge groß, und für Verfahren, die möglichst die ganze Szene dreidimensional erfassen müssen, um z. B. Modelle einer Person darin zu orten, klein sein sollte. Die maximal mögliche Basislänge von 24 cm der von mir verwendeten Stereokamera (Nr. 4) wurde daher für die Bestimmung des Kopfabstandes in unseren Versuchen an Treppen gewählt (siehe Abschnitt 3.4.2).

Insbesondere für statische Szenen kann statt der Nutzung einer Stereokamera eine monokulare Kamera bewegt und zeitlich nacheinander aufgenommene Bilder unterschiedlicher Perspektiven wie Bilder einer Stereokamera verwendet werden [264] (z. B. bei der Photogrammetrie in der Fernerkundung). Umgekehrt ist dies ebenfalls möglich, wenn eine statische monokulare Kamera ein sich bewegendes Objekt erfasst, wobei dies für die Personendetektion nicht nutzbar ist, sondern nur für statische Szenen auf einem Drehteller Anwendung findet.

Eine Stereokamera besitzt alle Möglichkeiten einer monokularen Kamera, da beide Perspektiven ein vollwertiges Bild erfassen, hat darüber hinaus aber viele Vorteile aufgrund der extrahierbaren 3D-Information:

- mächtiges Werkzeug zur Hintergrundsubtraktion, wenn sich die Kameraentfernungen von Vorder- und Hintergrund unterscheiden [195, 265–269],
- die Ermittlung der 3D-Information ist abgesehen von spiegelnden Reflexionen unabhängig von der Beleuchtung und auch den resultierenden Schatten [270],
- die Segmentierung sich teilweise verdeckender Objekte ist einfacher möglich,
- die reale Größe eines Objektes statt der Größe in der Bildebene, die 3D-Form oder der zurückgelegte Weg können, z. B. als zusätzliche Merkmale des Objektes, ermittelt werden.

Aufgrund der hinzukommenden Vorteile und der für die im Folgenden vorgestellten Verfahren robusteren Trefferquoten und besseren Fehlerraten habe ich mich bei der Entwicklung einer Methode zur markerlosen Personendetektion für die Nutzung einer Stereokamera entschieden (siehe Abschnitt 3.5). Die wichtigste Voraussetzung für eine robuste Gewinnung der 3D-Information aus Bildern einer Stereokamera ist neben einer ausreichenden Beleuchtung eine texturierte Oberfläche (siehe Abschnitt 3.4.3), um Punktkorrespondenzen in beiden Ansichten finden zu können.

Viele Verfahren setzen bei der Erkennung nicht nur die 3D-Information ein, sondern kombinieren diese mit Verfahren in der Bildebene. Dies ist einfach möglich, da die Tiefeninformation für einzelne Pixel bestimmt wird und damit diesen direkt zugeordnet werden kann und nicht von einer unabhängigen Einheit wie einem Laserscanner herrührt. In [269, 271] wird die Tiefeninformation zur Segmentierung von Bereichen gleicher Entfernung zur Kamera genutzt und anschließend die Klecks-Größe und -Form zur Detektion herangezogen. Für hohe Personendichten, wie sie bei unseren Versuchen vorkommen, kann die reine Klecks-Analyse nicht verwendet werden, da der Hintergrund oftmals gar nicht mehr zu sehen ist.

Ebenfalls die in [209, 272–274] vorgestellten Methoden bedürfen einer sauberen Segmentierung der einzelnen Vordergrundobjekte auch bei teilweise verdeckten Personen, was für dichte Menschenmengen wie bei unseren Experimenten schwer möglich ist. Andere Verfahren nutzen die Hautfarbe [266],

eine komplexere Gesichtserkennung [275] oder Bewegungsmuster der Arme und Beine [276, 277] zur Lokalisierung einer Person, was für die Überkopfansicht wie in unserem Versuchsaufbau nicht anwendbar ist, da insbesondere in dichten Situationen kaum Gesichter oder Extremitäten erkennbar sind. Eine strikte Seitenansicht ist für Verfahren nötig, die Bereiche gleicher Entfernung zur Kamera einer Person zuordnen [265, 278].

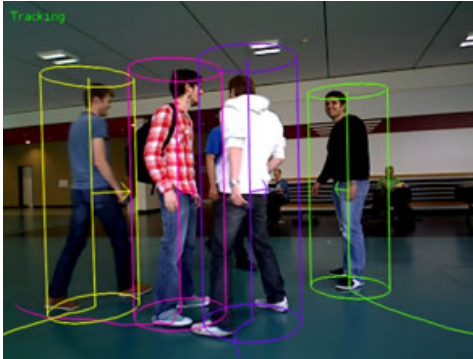


Abb. 2.9: Bestimmung der Position in der Bewegungsebene und Beschreibung des Aufenthaltsortes der Personen mittels Zylinder; Linien zeigen den Laufweg der Personen in der Bewegungsebene; Bild aus [34]

Wird die Grundfläche (*groundplane*) bestimmt, so kann die Position einer Person in der Bewegungsebene und der Kontaktpunkt mit dieser bestimmt werden [279, 280]. Dadurch kann der Suchraum möglicher Erscheinungen eines Menschen auch für beliebig ausgerichtete Stereokameras massiv reduziert werden, da mit dem Wissen über die Grundfläche bekannt ist, in welcher Ebene sich die Personen bewegen und damit die Lage der Ausprägungen des menschlichen Abbilds sich stark einschränken lässt [281] (siehe Abb. 2.9).

Wie bei den monokularen Kameras können auch durch überkopf nach unten ausgerichtete Stereokameras die Verdeckung in einem eingeschränkten Blickwinkel möglichst gering gehalten werden. Aus diesem Grund hatte auch ich die Perspektive entsprechend gewählt. Andere Verfahren

ren, die auch von solch einer Sicht ausgehen, sind in [282–290] beschrieben. In [284] wird als Merkmal die Anzahl der auf die Grundfläche projizierten 3D-Punkte herangezogen. Zur Bestimmung der Anzahl pro Flächeneinheit wurde dabei die Grundfläche in ein regelmäßiges Gitter für ein 2D-Histogramm aufgeteilt. Aufgrund der senkrechten Sicht ist die Anzahl von 3D-Punkten zur Merkmalsbestimmung für gewöhnlich ausreichend, die bei schräger Sicht pro Person je nach Entfernung wesentlich geringer ausfällt.

Eine Aufteilung der Bewegungsebene bzw. Grundfläche zur Erstellung eines 2D-Histogramms nutzen auch [34, 291, 292] für schräge Kameraansichten, um die Bereiche festzulegen, in denen intensiver nach Personen gesucht werden soll. Um der für größere Entfernungen geringer werdenden Anzahl an Punkten gerecht zu werden, wird jedes Datum mit dem Quadratabstand zur Kamera gewichtet. In Abb. 2.10 a ist zu erkennen, wie im nächsten Schritt gemäß einem Schnellverlagerungs-Algorithmus (*quick shift algorithm*) aus [293] Bereiche einzelner Objekte segmentiert werden. Zur besseren Detektion entfernter, durch wenige 3D-Punkte abgedeckter Personen hat [294] 3D-Volumenelemente eingeführt, die die 3D-Punkte aufnehmen und anschließend für das 2D-Histogramm verwendet werden. In [295] wird zur verlässlicheren Detektion insbesondere entfernter Personen das 2D-Histogramm nicht wie bisher mit Anzahlen von 3D-Punkten belegt, sondern mit der maximalen Höhe über dem Boden, was zu einem quantisierten Höhenfeld bzw. orthographischen Profil führt. [296], basierend auf [297], verzichtet ganz auf eine Quantisierung des Höhenfeldes und sucht mit einem bereichverschmelzenden Ansatz (*region growing*) zu einer Person gehörende 3D-Punkte und orientiert sich dabei an durchschnittlichen Abmessungen einer Person.

In meinem in Abschnitt 3.5 vorgestellten Verfahren zur markerlosen Personendetektion verzichte ich auf eine aufwändige Bewertung der in das Weltkoordinatensystem rückprojizierten Disparitätswerte

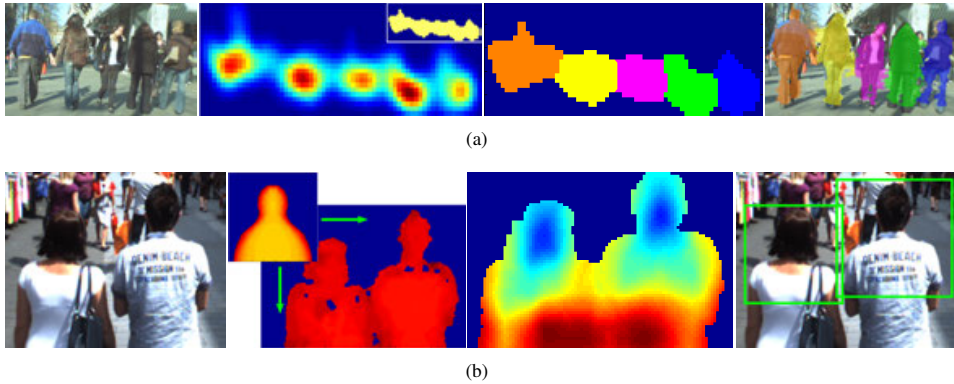


Abb. 2.10: (a) 2D-Histogramm der 3D-Punkte in der Bewegungsebene mit anschließender Ermittlung von Häufungen [293] zur Lokalisierung und Separation von einzelnen Personen; (b) Detektion zweier Oberkörper durch Vorlagen-Abgleich im perspektivischen Tiefenfeld mit einem sich verschiebenden Fenster [298]; das dritte Bild zeigt die Abstandsmatrix zwischen Vorlage und dem Tiefenfeld der realen Szene (blau geringer, rot großer Abstand); Bilder aus [34]

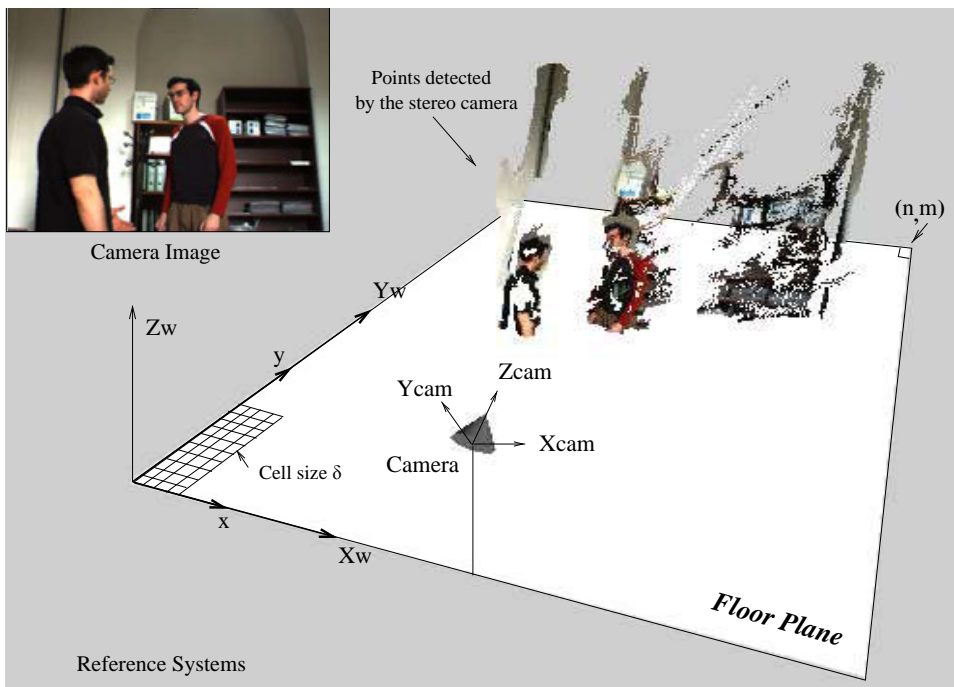


Abb. 2.11: Pixel des oberen linken Kamerabildes projiziert in das Weltkoordinatensystem; Zellen auf der Grundfläche deuten die Diskretisierung der Fläche für ein 2D-Histogramm an; Bild aus [275]

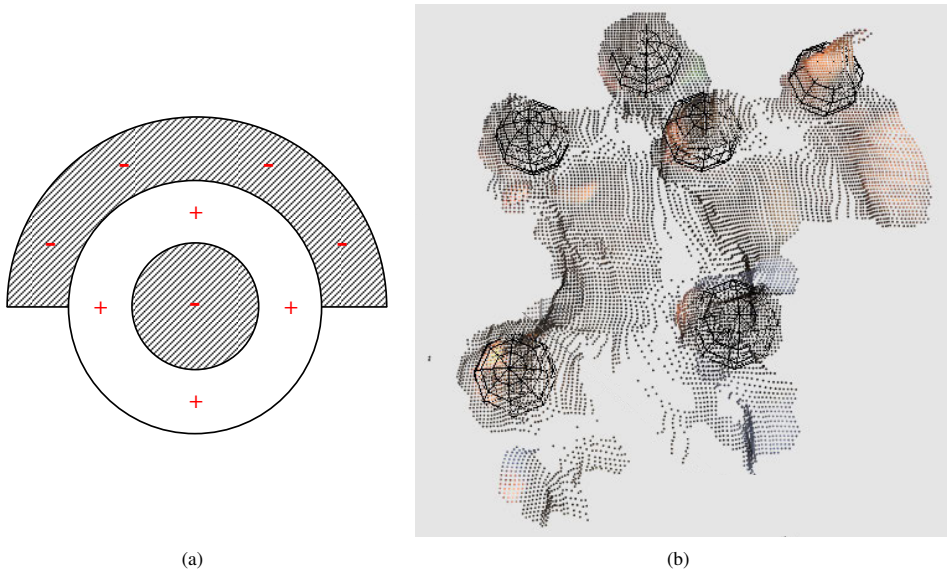


Abb. 2.12: Suche nach Kugelstrukturen in einer aus der Disparitätsmatrix ermittelten Punktwolke: (a) Querschnitt der Vorlage mit negativ und positiv gewichteten Orten, bei der der positive Bereich das Spektrum der möglichen Kopfgrößen beinhaltet; (b) Punktwolke mit sechs eingebetteten Kugel, an denen Köpfe detektiert wurden; basiert auf Bildern aus [286]

(*plan-view statistics*), im Gegensatz zu [275, 284, 294–296, 299–302], die ihre Erkennungsmerkmale aus den umgerechneten 3D-Punkten gewinnen (siehe Abb. 2.11, vgl. Abb. 3.45). Der Verzicht auf die Rückprojektion ist nur möglich, da von einer festen Perspektive ausgegangen wird, wie es auch [34] für den Nahbereich tut. Die für den Nahbereich verwendete Methode basiert auf Arbeiten aus [298] zur effizienten Gewinnung von Aufenthaltsorten von Personen für eine Stereokamera-Sicht etwa auf Augenhöhe der zu detektierenden Personen, wie es für mobile Roboter oder Helmkameras der Fall ist. Abb. 2.10 b zeigt, wie eine Vorlage (*template*) von einem perspektivischen Tiefenfeld eines Oberkörpers mit einem sich verschiebenden Fenster (*sliding window*) direkt mit dem perspektivischen Tiefenfeld der zu untersuchenden Szene verglichen wird. Im dritten Bild ist die Abstandsmatrix farb-kodiert dargestellt (blau geringer, rot großer Abstand).

In [286] wird aufsetzend auf der Arbeit [304] in der aus dem überkopf aufgenommenen Tiefenfeld gewonnenen Punktwolke nach Orten gesucht, die in Form einer der Größe des menschlichen Kopfes entsprechenden Kugel angeordnet sind, wobei Punkte neben dieser Kugel negativ in die Bewertung einer möglichen Kopfposition einfließen (siehe Abb. 2.12). Die dokumentierte Richtig-Positiv-Rate von 88% in jedem Einzelbild und 96% einschließlich eines verbindenden Trackings für eng beieinander laufenden Personen ist hoch (vgl. Tab. 3.11).

Dem in dieser Arbeit vorgestellten Verfahren zur markerlosen Personendetektion mittels einer Stereokamera kommt das später veröffentlichte Verfahren in [285] am nächsten, das nach lokalen Minima im Abstands- bzw. Tiefenfeld einer überkopf senkrecht nach unten ausgerichteten Stereokamera sucht, um Köpfe von Personen zu lokalisieren. Dabei wird kein weiteres Wissen von Bildinformationen wie z. B. Farben und über Standorte in vorangegangenen Einzelbildern herangezogen. Folgende Schritte werden hierfür nacheinander genannt (siehe Abb. 2.13):



Abb. 2.13: Vorverarbeitungsschritte des Tiefenfeldes zur anschließenden Leitkegelsuche: Originalbild, perspektivisches Tiefenfeld mit weißen Punkten an Stellen initialer Leitkegel und vier gelben Punkten an Positionen lokaler Minima, orthografische Projektion des perspektivischen Tiefenfeldes, Anwendung eines dünn besetzten Medianfilters mit einem großen Kern; Bilder aus [285]

1. Auswahl des Abstandsbereichs aus dem perspektivischen Tiefenfeld, der dem menschlichen Körper entspricht,
2. Anwendung eines Medianfilters auf das perspektivische Tiefenfeld, um Rauschen bzw. Fehler zu bereinigen,
3. orthografische Projektion des perspektivischen Tiefenfeldes auf die Grundfläche bzw. Bewegungsebene zur Überführung in ein reguläres Tiefenfeld,
4. Anwendung eines dünn besetzten Medianfilters mit einem großen strukturierenden Element bzw. Kern (*kernel*) zur Förderung einer konvexen Struktur des Tiefenfeldes und
5. Suche nach lokalen Minima im Tiefenfeld.

Zur Suche nach lokalen Minima wird das Tiefenfeld entsprechend der erwarteten Kopfgröße unterabtastet (*downsampling* oder *subsampling*) und die Punkte, als Leitkegel bezeichnet, herausgesucht, deren direkte vier Nachbarn entfernter zur Kamera liegen. Gruppen von erhöht liegenden Leitkegeln werden zusammengefasst und als Kopfposition der Mittelwert der Positionen der Leitkegel verwendet. Die Suche nach Leitkegeln im orthografisch projizierten Tiefenfeld wendet [305] auf Schrägsichten an.

Aufgrund von Ungenauigkeiten in der zu ermittelnden Disparitäts-Matrix insbesondere für ferne oftmals kleine Objekte ist bisher auch für Stereokameras noch kein Verfahren gefunden worden, das für alle Szenarien verlässlich Personen detektiert, wie es dem Menschen möglich ist. Das selbst entwickelte Verfahren (siehe Abschnitt 3.5) zeigt abgestimmt auf unsere Bedürfnisse sehr gute Ergebnisse (siehe Abschnitt 3.6.6).

2.3.3 Mehrkamarasystem

Von Mehrkamarasystemen spricht man bei mehr als zwei zusammen arbeitenden monokularen Kameras oder zwei Kameras, die nicht für die Erstellung eines gemeinsamen Tiefenfeldes genutzt werden. Für sich bewegende Objekte wie Personen sind für die gemeinsame Erhebung von Laufwegen zeitlich synchronisierte Kameras nötig (siehe Abschnitt 4.1.1.8).

Mehrkamarasysteme bieten eine robustere Bestimmung des 3D-Ortes als Stereokameras und besitzen neben den Möglichkeiten einer Stereokamera die Möglichkeit einer Abdeckung größerer Gebiete u. a. durch nur teilweise überlappende Perspektiven. Zudem können Verdeckungen verringert und Objekte ohne Bewegung komplett dreidimensional erfasst werden. Gerade für eine überschaubare Anzahl von

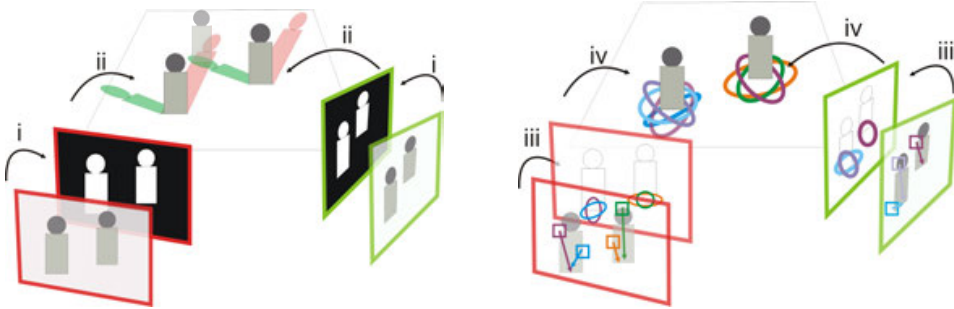


Abb. 2.14: (links) klassische Methode zur Bestimmung der Position einer Person (i) durch Projektion des Vordergrundes (ii) auf die Bewegungsebene; eine Person wird dort angenommen, wo die Projektionen der Vordergründe sich überlappen; dabei können, wie hier halbdurchsichtig visualisiert, zwischen den Personen Fehldetektionen auftreten (*ghosting*); (rechts) (iii) Hough-Transformation jeder Perspektive zur Bestimmung möglicher Positionen; (iv) eine Person wird dort angenommen, wo die Projektionen der Hough-Transformation sich überlappen und damit die Position exakter eingrenzen; basiert auf Bildern aus [310]

sich temporär verdeckenden Personen, die über längere Zeit stabil verfolgt werden sollen (z. B. bei Sportveranstaltungen), bieten Mehrkamerasysteme eine robuste Extraktion von Laufwegen. Die Option zur Erweiterung des einzusehenden Bereichs wurde auch bei den meisten unserer Versuche durch mehrere sich zum Teil überlappende Kameras oder Stereokameras genutzt (siehe Abschnitt 4.1.1.2).

Kamerasysteme mit mehr als zwei Kameras erlauben, z. B. wie bei der von mir eingesetzten Kamera Nr. 4 (siehe Tab. 4.2), die Basislänge spontan zu wechseln, um den optimalen Abstandsbereich wählen zu können. Sind Kameras nicht entlang einer Linie angeordnet (z. B. [308]), können Probleme bei der Bestimmung des Kameraabstandes durch gleichförmige oder sich wiederholende Objekte entlang von Pixelzeilen vermieden werden (siehe Abschnitt 3.4.3).

Auch im Bereich der Überwachung werden Mehrkamerasysteme eingesetzt, u. a. um einmal erfasste Individuen über die zum Verbund gehörenden Kameras hinweg wiederzuerkennen und zu verfolgen. In [309] werden mehrere Kameras eingesetzt, um Personen zu lokalisieren, zu verfolgen und zu identifizieren. Zur Lokalisierung werden die aus den verschiedenen Kameraperspektiven über ein Hintergrund-Modell, basierend auf zusammengesetzten Gauß-Verteilungen, ermittelten Vordergründe auf die gemeinsame Bewegungsebene projiziert, deren Schnittmengen die Positionen der Personen angeben (siehe Abb. 2.14 links). Weitere Merkmale auf Pixelebene werden zur genaueren Detektion einer Person herangezogen. Für die Verfolgung und Wiedererkennung wird ein 3D-Modell jeder Person erstellt, das u. a. die Personengröße beinhaltet. Die Personengröße wird mittels Projektion des Vordergrundes auf zur Grundfläche parallele Ebenen ermittelt. Beim korrekten Ebenenabstand berühren sich gerade die aufeinandergelegten Projektionen auf die Grundfläche und auf die Ebene in Kopfhöhe. Die Orientierung der Person wird aus dem bisher erfassten Laufweg geschätzt.

In [310] wird zur Bestimmung der Position in der Bewegungsebene nicht der häufig verwendete Überlappungsbereich des auf die Grundfläche projizierten Vordergrundes aller Kameras gewählt, da hierbei von nah beieinander stehenden Personen Fehldetektionen hervorgerufen werden können (*ghosting*) (siehe Abb. 2.14). Stattdessen wird eine Überlappung von Hough-Wertungen nach der Anwendung von Hough-Transformationen der einzelnen Perspektiven [311, 312] durchgeführt, um Fehldetektionen zu minimieren.

[280, 313] benutzen sequentielle Monte-Carlo-Methoden bzw. Partikel-Filter (*particle filter*) zum Ver-



Abb. 2.15: Verzerzte Kameraperspektiven, bei denen Pixel in Kopfhöhe auf die gleiche Pixelkoordinate fallen; Pixel mit der geringsten Varianz sind Kopfpositionen; Bilder aus [314]

folgen von Personen über mehrere Kameras hinweg.

Zur Bestimmung der Kopfposition werden die Bilder der einzelnen Kameraperspektiven in [314] so verzerrt, dass die Resultate die Köpfe in Kopfhöhe auf dieselben Pixel abbilden und damit die Pixel mit der geringsten Varianz die Kopfpositionen markieren (siehe Abb. 2.15).

Um den Gang einer Person zu erfassen, wird in [315] ein 3D-Voxel-Modell aus Silhouetten mehrerer Kameraperspektiven abgeleitet.

Von den in Tab. 2.1 aufgeführten Versuchen wurden bei [143] aufgrund geringer Deckenhöhe zur exakten Lokalisierung die auf den Schultern eines weißen T-Shirts angebrachten farbigen Tischtennisbälle mit drei Kameras aufgezeichnet (siehe Abb. 2.7 e). Durch die Anzahl der Kameras wurde gewährleistet, dass mindestens zwei Kameras die beiden Tischtennisbälle während des gesamten Versuchs sehen. Mit einem Verfahren basierend auf [316] wurde die 3D-Position der Person als Mittelpunkt der beiden farbigen Marker bestimmt. Die Bälle sind wesentlich kleiner als die farbigen Kopfbedeckungen der anderen in Abschnitt 2.3.1.5 vorgestellten Versuche und erlauben damit auch bei unterschiedlichen Perspektiven eine präzisere Lokalisierung.

Weitere Verfahren zur Detektion von Personen in Mehrkamerasystemen sind in [181, 208, 271, 274, 302, 317–321] aufgeführt.

2.3.4 Infrarotkamera

Infrarotkameras erfassen elektromagnetische Wellen im Spektralbereich zwischen sichtbarem Licht und der längerwelligen Terahertzstrahlung (Wellenlängen zwischen $7,8 \cdot 10^{-7} \text{ m}$ und 10^{-3} m). Sensoren herkömmlicher Kameras sind oft in der Lage einen Teil dieser Wellenlängen zu erfassen. Diese werden aber für gewöhnlich durch Infrarot-Sperrfilter zurückgehalten.

Bei der Personendetektion muss zwischen Techniken und Verfahren unterschieden werden, die nahes (Wellenlängen zwischen $7,8 \cdot 10^{-7} \text{ m}$ und $3 \cdot 10^{-6} \text{ m}$) oder mittleres Infrarot (thermische Strahlung, Wellenlängen zwischen $3 \cdot 10^{-6} \text{ m}$ und $5 \cdot 10^{-5} \text{ m}$) erfassen bzw. nutzen. Die mittleres Infrarot empfangenden Wärmebildkameras werden separat in Abschnitt 2.3.5 behandelt.

Nahes Infrarot wird zur Erkennung von Personenbewegungen auf verschiedene Arten angewandt. Dabei wird immer ausgenutzt, dass Infrarotlicht durch das menschliche Auge nicht wahrgenommen werden kann.

In dunkler Umgebung ist es für ein Überwachungssystem aufgrund aktiver Aussendung von Infrarotlicht möglich, unbemerkt ein Bild der Umgebung zu machen. Das erfasste Bild der Intensität des Infrarotlichtes hat Ähnlichkeit mit Grauwertbildern, besitzt aber nicht die gleiche Helligkeitsverteilung wie Bilder von Kameras, die vom Menschen wahrnehmbare elektromagnetische Strahlung erfassen.

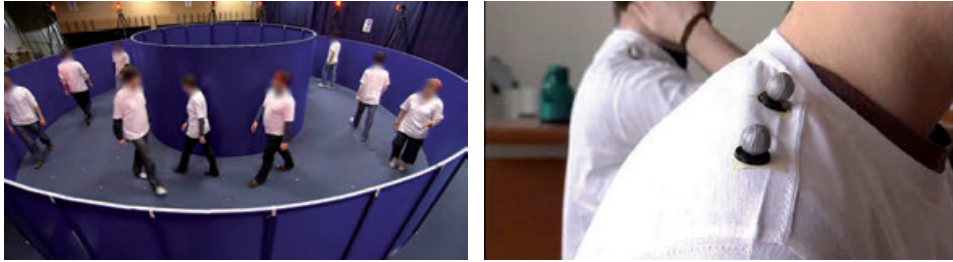


Abb. 2.16: Versuchsaufbau mit blauen Wänden und umliegenden Infrarotkameras und zwei auf einer Schulter befestigten Infrarotmarkern beim Pedigree-Projekt; Bilder aus [325]

Trotzdem können ähnliche Verfahren wie für herkömmliche Kameras aus Abschnitt 2.3.1 verwendet werden [322, 323]. Eine Übersicht von Verfahren kann [324] entnommen werden.

In heller Umgebung besteht die Möglichkeit, relevante Punkte mit besonders gut Infrarotlicht reflektierenden Markern zu versehen, um diese durch ausgestrahltes Infrarotlicht mittels Infrarotkameras leicht zu lokalisieren. Ein Beispiel ist Abb. 2.3 c zu entnehmen, in dem die Lage von für die Pose wesentlichen Punkten des Körpers aufgrund von Selbstverdeckungen mit einer Vielzahl von Infrarotkameras erfasst werden, um die Bewegung eines virtuellen Charakters zu steuern. Sogar die Gesichtsmimik kann über diese Methode detektiert werden. Die gleiche Technik wird in VR-Umgebungen verwendet, um die Position und Blickrichtung des Akteurs und die Lage seiner Eingabegeräte zu registrieren.

Diese Art von Markern kann auch nur zur Erfassung der Laufwege von Personen genutzt werden. In Versuchen der Tab. 2.1 wurden Infrarotlicht reflektierende Marker von [139, 155–157] verwendet.

In [139] wird diese Methode nur in dunkler Umgebung angewendet. Ansonsten wurden farbige Kopfbedeckungen genutzt. Dabei wurden nicht reflektierende Bälle, sondern aktiv Infrarotlicht aussendende LEDs am Kopf der Versuchsteilnehmer befestigt.

Aufwändiger wurde die Erfassung in den dieselbe Technik verwendenden Versuchsreihen aus [155–157] realisiert, die im Rahmen des Pedigree-Projektes (*PEDEstrlan GRoups: EmErgence of collective behaviour through experiments, modelling and simulation*) durchgeführt wurden [325].

Eingesetzt wurde dabei ein kommerzielles Tracking-System der Firma Vicon [327]. Vier Infrarotmarker wurden auf den Schultern und auf dem Kopf platziert und mithilfe von 12 Infrarotkameras erfasst. Der Mittelwert der vier Positionen aller Marker einer Person wurde als Standort verwendet.

Bei der Verwendung von Infrarotlicht nutzenden Verfahren ist zu beachten, dass Sonnenlicht ebenfalls Infrarotlicht enthält und somit Verfahren, die zeitlich oder örtlich strukturiertes Infrarotlicht benötigen, nicht bei Tageslicht verwendet werden können. Ein Beispiel hierfür ist die Tiefenermittlung durch die Kinect (siehe Abschnitt 2.3.6 und Abb. 2.3 b).

Die Nutzung der Infrarotmarker sehe ich insbesondere bei solchen Anwendungen als sinnvoll an, die sehr detaillierte Bewegungen in wohl definierter Umgebung erfassen sollen. In natürlichen Umgebungen wie dem Stadion in unserer Versuchsreihe Nr. 6 (siehe Tab. 4.1) werden reflektierende Objekte wie Geländer die Erkennung erschweren. Herkömmliche Kameras sind zudem günstiger als Infrarotkameras.

2.3.5 Wärmebildkamera

Eine Wärmebildkamera oder auch Thermografiekamera ist eine spezielle Infrarotkamera, die zur Messung von Wärmestrahlung nur mittleres Infrarot erfassen kann. Sie ist damit die erste der bisher vorgestellten Kameras, die nicht nur die von Objekten reflektierten elektromagnetischen Wellen detektiert, sondern insbesondere die Eigentemperatur der Objektoberfläche misst.



Abb. 2.17: Intensitätsbilder einer Wärmebildkamera einer Straße mit Passanten bei Nacht, bei Tag und bei Regen; Bilder aus [328]

Sie ermöglicht daher auch im Dunkeln ohne aktive Aussendung von Licht die Erfassung einer Szene und insbesondere von Personen. Daher werden Sensoren dieses Typs auch als PIR-Sensoren (*passive infrared*) bezeichnet. Abb. 2.17 zeigt Intensitätsbilder einer Wärmebildkamera einer Straße mit Passanten bei Nacht, bei Tag und bei Regen. In allen Situationen ist die Kontur der Personen zu erkennen und in der Nacht und bei Regen besser als mit den bisher vorgestellten Kameraarten [328]. Bei Dunkelheit und schlechtem Wetter ist die Nutzung einer Wärmebildkamera daher sinnvoll. Zudem liefert sie aus der Luft in unbewohnten Gebieten bei nicht zu starker Sonneneinstrahlung einen Hinweis auf Wärmequellen wie z. B. Personen.

Zur Detektion von Personen in Wärmebildern sind ähnliche Verfahren wie bei sichtbares Licht erfassenden Kameras nutzbar wie z. B. Verfahren, die mittels Klecks-Analyse (*blob analysis*) oder Vorlagen-Abgleich (*template matching*) arbeiten [49, 329–338]. Auch bei Wärmestrahlung können Reflexionen auftreten und Personen dadurch mehrfach erfasst werden, wie z. B. an Schaufensterscheiben (siehe auf dem Boden der rechten Abb. 2.17).

Zum Test, ob eine Wärmebildkamera bei dichten Menschenmengen überkopf Personen zuverlässig ohne Marker detektieren kann, wurde bei den Hermes-Versuchen (Nr. 6) neben den Stereokameras auch eine Wärmebildkamera, mit der optischen Achse senkrecht zum Boden ausgerichtet, etwa 6 m über den Köpfen der Teilnehmer aufgehängt. Die Idee war, dass der Kopf sich als besonders warmer ohne Kleidung bedeckter Körperteil vom restlichen Körper und der Umgebung abhebt. Die Abb. 2.18 zeigt, dass sich der Kopf durch Haare oder Kopfbedeckung eher als Kältebereich des Körpers darstellt. Dies ist aber stark abhängig von der körperlichen Verfassung, der

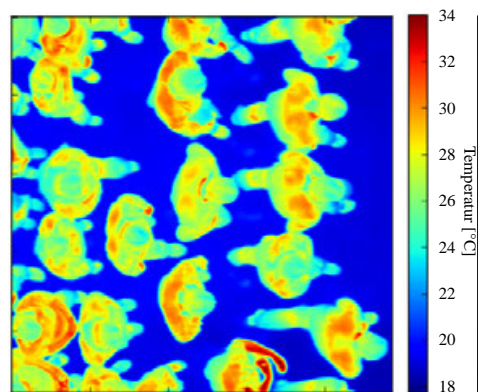


Abb. 2.18: Bild einer überkopf angebrachten Wärmebildkamera eines Hermes-Versuches (Nr. 6)

Behaarung und der getragenen Kleidung. Ist der Hintergrund um eine Person zu sehen, zeigt die Silhouette genau den Standort einer Person. Dies ist aber auch durch eine Hintergrundsubtraktion bei Bildern herkömmlicher Kameras möglich, wobei es hierbei keinen Schattenwurf, dafür aber neben Reflexionen Artefakte durch erwärmende Untergründe an Stellen, an denen eine Person gestanden hat, gibt. Nah zusammenstehende Personen oder schräge Sichten lassen die Bereiche mehrerer Personen verschmelzen. Da Wärmebildkameras teurer als herkömmliche Kameras und auch Stereokameras sind, wurde diese Art der Personenerkennung nicht weiter verfolgt.

Ein Vergleich verschiedener überkopf angebrachter Personenzählssysteme einschließlich Systemen auf Basis von Wärmebildkameras wird in [339] durchgeführt.

In den in [153] beschriebenen Versuchen (siehe Tab. 2.1) wird eine kommerzielle Software und abgestimmte Hardware der Firma Irisys verwendet (siehe Anhang E). Dabei kommt eine Wärmebildkamera zur reinen Zählung zum Einsatz.

2.3.6 3D-Kamera

Die Steigerung der Stabilität der Personendetektion durch Stereokameras aufgrund der hinzukommenden Tiefeninformation wurde bereits in Abschnitt 2.3.2 beschrieben.

Es gibt neben Stereokameras oder Mehrkamerasystemen eine Vielzahl weiterer Techniken, die es ebenfalls erlauben, die Tiefe bzw. die Entfernung zum Sensor von einer Szene zu bestimmen. Dabei unterscheiden sich die Methoden u. a. in der Tiefenauflösung (Genauigkeit der Tiefeninformation ggf. abhängig von der Entfernung zum Sensor) und dem Tiefenbereich (zwischen minimal und maximal bestimmbarer Tiefe) [340]. Die sehr speziellen Techniken der Interferometrie, der Kohärenzlängenmessung oder der Schattierungsanalyse [341] sind für den Zweck der Entfernungsbestimmung von sich bewegenden Menschen ungeeignet, da sie für sehr kleine detaillierte oder großskalige Messungen sowie statische Objekte eingesetzt werden.

Die restlichen Verfahren lassen sich aufteilen in Methoden, die durch Triangulation die Entfernung bestimmen, wie dies die Stereokamera tut (vgl. Abb. 3.41), oder die aufgrund der Laufzeit (*time-of-flight*, TOF) eines Signals wie bei einem Laserscanner die Entfernung messen (siehe Abschnitt 2.3.8 und 2.3.9).

Triangulationsbasierte Tiefenmesstechniken nutzen die Unterschiede eines Bildes aus verschiedenen Blickwinkeln. Hierzu gehören neben der Stereoskopie und den für unsere Anwendung ungeeigneten Verfahren mittels Fokussierung und Schärfe (z. B. durch Lichtfeldkameras bzw. plenoptischen Systemen) oder konfokalen Mikroskopie auch aktive Triangulationsmethoden, die eines der beiden benötigten Bilder selbst erzeugen und statt zwei Bilder zu erfassen, das selbst erzeugte Bild abstrahlen und dieses von einer anderen Perspektive erfassen. Alle Triangulationsmethoden besitzen eine quadratisch mit der Entfernung zur Kamera abfallende Tiefenauflösung (vgl. Abschnitt 3.4.3).

Bei der Erfassung von Personen dürfen diese nicht durch das Messsystem beeinflusst oder gefährdet werden, daher können u. a. Verfahren, die für die aktive Triangulation für den Menschen wahrnehmbare Muster zur Bestimmung der Tiefeninformation projizieren [342] nicht herangezogen werden. Bei folgendem System wird daher mit einem Muster im Infrarotspektrum gearbeitet.

2.3.7 RGB-D-Sensor

Ein aktuell sehr verbreitetes System, das mit aktiver Triangulation arbeitet, ist die Microsoft Kinect [92] oder Asus Wavi Xtion. Die Gründe für den hohen Verbreitungsgrad [343] sind, neben der Einsatzmöglichkeit im Spielektor, die im Gegensatz zu anderen 3D-Kameras für die erzielbare Tiefenauflösung sehr geringen Anschaffungskosten sowie eine darauf abgestimmte freie und umfangreiche Softwarebibliothek [344], die neben der 3D-Information z. B. bereits die Lage auch mehrerer menschlicher Körper orten (siehe Abb. 2.3 b) und Gesten in Seitenansichten ohne Einsatz von Eingabegeräten oder Markern erkennen kann.

Das aktiv ausgesendete bzw. projizierte Bild der Kinect besteht aus einem wiederholungsfreien Punktmuster im infraroten Lichtspektrum, so dass es vom menschlichen Auge nicht wahrgenommen werden kann. Dabei bietet die Kinect zusätzlich eine integrierte Kamera, deren Pixel die Tiefeninformation zugeordnet werden können, weshalb bei dieser Art von Sensoren auch von RGB-D-Sensoren (*red-green-blue-depth*) gesprochen wird. Die Tiefenauflösung ist vergleichbar mit der einer Stereokamera [345,346]. Damit sind Verfahren für Stereokameras (siehe Abschnitt 2.3.2) auch für RGB-D-Sensoren anwendbar [34]. Als Nachteil gegenüber anderen Systemen und eine Einschränkung für den Einsatz zur Detektion von Personen ist der geringe Tiefenbereich zwischen 0,8 m und 4 m und aufgrund der Nutzung eines Infrarotlicht-Musters der nicht mögliche Einsatz im Sonnenlicht. Die Systeme sind insbesondere zur Personenerkennung von der Seite [347–351] und Erfassung der Bewegung bzw. Gesten entwickelt worden [95, 352, 353]. Die gesamte Erkennung arbeitet in Echtzeit bei einer Bildwiederholrate von 30f/s und erlaubt mit einer ebenfalls in Echtzeit durchführbaren simultanen Lokalisierung und Kartenerstellung (*simultaneous localization and mapping*, SLAM) die 3D-Erfassung des gesamten Körpers einer Personen [354] und beim Mitführen des RGB-D-Sensors die Lageschätzung (Position und Orientierung), die in der Robotik eingesetzt wird [355] (Odometrie), aber auch für die Erfassung eines Laufweges einer Person anwendbar wäre. Die simultane Lokalisierung und Kartenerstellung geschieht dabei durch eine kontinuierliche Punktzuordnung aufeinanderfolgender Aufnahmen [350, 356, 357], z. B. mittels der Suche nach Übereinstimmung einer zufälligen Stichprobe (*random sample consensus*, RANSAC) oder einer sukzessiven Anpassung der Koordinatentransformation eines Datensatzes zur Minimierung der Punktabstände (*iterative closest point*, ICP) [358]. Beim Einsatz zur Erfassung eines Laufweges kann es aufgrund des Verzichtes auf externe Messsysteme oder Ortungssignale (vgl. inertielle Navigationssysteme in Abschnitt 2.3.10) zum Drift in der Trajektorie kommen.

Um mehrere Personen in dichten Menschenmengen zu verfolgen, muss auch die Kinect überkopf nach unten ausgerichtet sein (z. B. in [359–362]). In [359] wurden in diesem Sinne mehrere sich in Kopfhöhe leicht überlappende Kinect-Systeme an der Decke eines Flures befestigt. Abb. 2.19 zeigt aus dieser Perspektive ein Tiefenfeld bzw. eine Tiefenkarte und eine Punktwolke, die aus einer Tiefenkarte durch Rückprojektion extrahiert wurde. Beim vorgestellten Verfahren wird zur Detektion der Personen zuerst eine Hintergrundsabstraktion durchgeführt und die verbleibenden Punkte auf den Bereich des Oberkörpers beschränkt, indem Punkte vom Boden gesehen oberhalb von 2,1 m und unterhalb von 1,5 m verworfen werden. Anschließend werden Gruppen von Anhäufungen der Punkte mittels der agglomerativen hierarchischen Gruppenanalyse mit maximalem Abstand aller Elementpaare für alle Gruppen (*complete-linkage clustering algorithm*) [363] mit einer maximal schulterbreiten Ausdehnung von 0,6 m gesucht. Dafür wird zur Rechenzeiterparnis zufällig eine Untermenge der Punktwolke herangezogen. Kleine Punktgruppen werden danach verworfen. Innerhalb der Punktgruppen wird der Punkt als Position gewählt, dessen Bodenabstand im 95%-Perzentil der Abstände liegt. Die Richtig-Positiv-Rate, also der Prozentsatz der erkannten Personen, wird mit ca. 94% und der auch

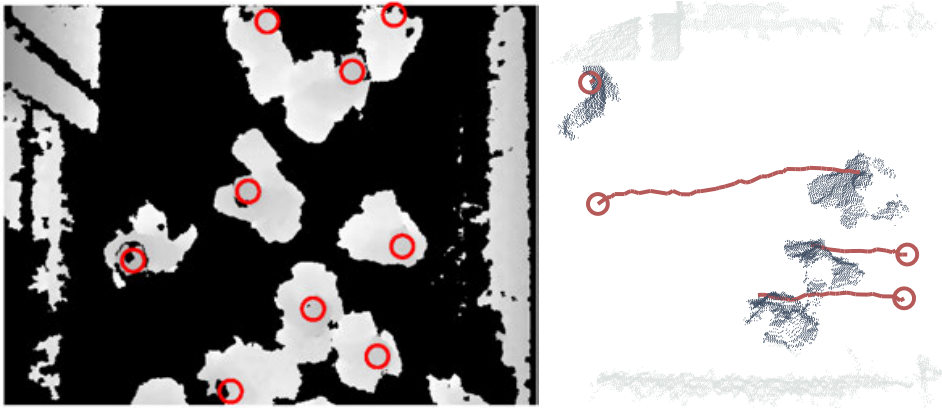


Abb. 2.19: (links) perspektivisches Tiefenfeld (vgl. Abb. 3.50 b) einer überkopf nach unten auf Personen gerichteten Kinect als Grauwertbild mit manuell markierten Kopfpositionen; (rechts) rückprojizierte Punktwolke des auf den Oberkörper eingeschränkten Tiefenfeldes mit Trajektorien detektierter Personen; Bilder aus [359]

von uns verwendete Positionsfehler gemäß der MOTP-Metrik (*multiple object tracking precision metric*) [364] (vgl. Gleichung 3.40) mit etwa 4 cm angegeben. Damit ist die Richtig-Positiv-Rate größer (vgl. Abschnitt 3.6.6.1) und der Positionsfehler ähnlich (vgl. Abschnitt 3.6.6.2) dem in dieser Arbeit vorgestellten markerlosen Verfahren (siehe Abschnitt 3.5). In [360, 361] wurde dieses Verfahren [359] zur Modellbildung herangezogen.

Für die Erweiterung des eingeschränkten Beobachtungsbereiches wurden in [359] mehrere überlappende Kinects eingesetzt. Die Qualität der Tiefenkarte sinkt im Überlappungsbereich durch sich überlagernde Muster. Unbestimmbare Tiefen erscheinen als Löcher im perspektivischen Tiefenfeld [365–367]. Für die Abdeckung der Versuchsfläche zur Erfassung aller Laufwege in größeren Versuchen (siehe Kapitel 4) sind mehrere Kinects nötig. In [368, 369] werden zur Lösung der Problematik synchronisierte Belichtungen bzw. Verschlüsse (*shutter*) vorgestellt, was zu einer Verringerung der Bildwiederholrate führt. Bei [370] wird der Überlappungsbereich möglichst gering gehalten, was jedoch nur für eine feste Entfernung zum Sensor funktioniert und für den Fall der Laufwegebestimmung überkopf Personen je nach Größe unterschiedlich stark betroffen wären. Die generell anwendbare und robusteste Methode wird in [371, 372] beschrieben, bei der ein Vibrationsmotor an der Kinect dafür sorgt, dass das Muster richtungsabhängig verschmiert und somit zu unterscheidbaren Mustern führt. Jedoch führt dies auch dazu, dass das Farbbild unscharf wird. Um Löcher zu schließen, die unter anderem durch überlappende Kinects entstanden sind, wird in [373] ein angepasster Medianfilter vorgestellt. Für überlappende Sichtfelder besitzen passive Techniken wie Stereokameras einen Vorteil, da sie nicht die Umgebung beeinflussen, sondern nur passiv messen (vgl. Abschnitt 4.1.1.2).

In [362] wird ebenfalls die Personendetektion überkopf durchgeführt. Dabei wird das Tiefenfeld durch morphologische Operationen stark geglättet und anschließend mittels Gradientenverfahren (*gradient fill algorithm*), ausgehend von den gruppierten lokalen Minima, die Schwerpunkte der Regionen, die die oberen Kopfbereiche beschreiben, bestimmt, die damit die Positionen der Personen kennzeichnen (siehe Abb. 2.20). Diese Methode bewirkt das Ähnliche wie das Heranziehen des zweitobersten Ellipsenmittelpunktes zur Festsetzung der Position einer Person im in dieser Arbeit vorgestellten markerlosen Verfahren (siehe Abschnitt 3.5.5).

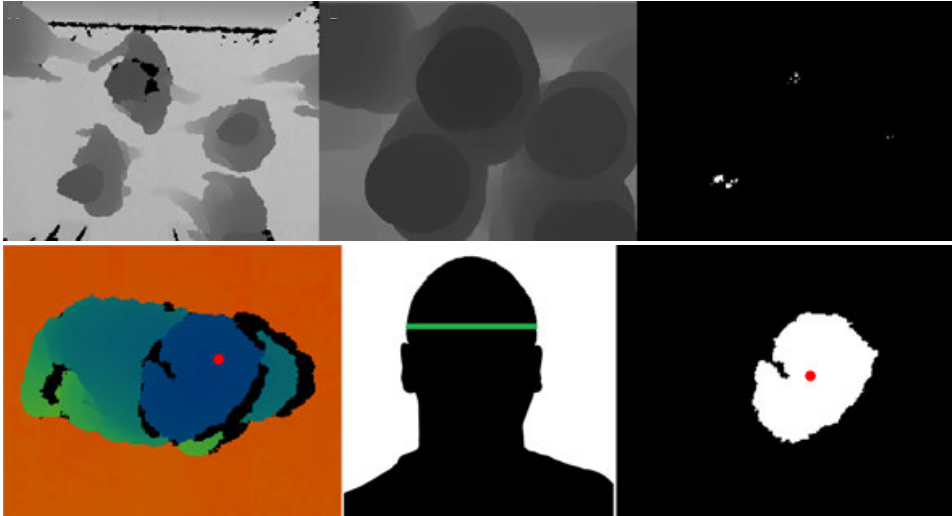


Abb. 2.20: Schritte zur Detektion der Kopfmitte aus perspektivischen Tiefenfeldern einer Kinect; von links oben nach rechts unten: original perspektivisches Tiefenfeld der Kinect; geglättet durch Dilatation mittels eines kreisförmigen schulterbreiten Strukturelementes; weiße Punkte zeigen lokale Minima im Tiefenfeld; roter Punkt im farbkodierten Tiefenfeld ist Mittelpunkt gruppierter Minima; grüne Linie zeigt den im Folgenden einzugrenzenden Höhenbereich; roter Punkt ist Schwerpunkt des vom lokalen Minimum ausgehenden mittels Gradientenverfahren ermittelten Höhenbereichs des perspektivischen Tiefenfeldes; Bilder aus [362]

Auch mit der Kinect werden viele Verfahren vorgestellt, die neben dem Tiefenfeld das ebenfalls von der Kinect stammende Farbbild oder weitere Daten von anderen Sensoren gemeinsam nutzen (*sensor fusion*), um Personen zu detektieren [376–380].

Aufgrund der vorgestellten Möglichkeiten bieten sich die Kinect oder vergleichbare RGB-D-Sensoren an, um Personen wie mit einer Stereokamera markerlos zu erkennen und zu verfolgen. Hierbei liegt die Stärke der Kinect in der schnellen, kostengünstigen und robusten Erstellung eines Tiefenfeldes, ohne vom Umgebungslicht und von der Textur der Umgebung abhängig zu sein. Aufgrund der Nutzung von Infrarotlicht-Mustern und dem eingeschränkten Tiefenbereich ist die Verwendung auf Innenräume beschränkt. Dieser Einsatz soll in Zukunft getestet (siehe Abschnitt 5.2) und die Anwendbarkeit der für Tiefenfelder einer Stereokamera entwickelten markerlosen Erkennung geprüft werden.

2.3.8 Laufzeit-Kamera

Neben den auf Triangulation basierenden bisher vorgestellten Techniken zur Tiefen- bzw. Abstandsbestimmung nutzt die zweite große Gruppe von Methoden die Laufzeit von Signalen. Der Abstand z lässt sich dabei einfach aus der Laufzeit t und der Ausbreitungsgeschwindigkeit v des Signals bestimmen:

$$z = \frac{t \cdot v}{2}.$$

Diese Formel zeigt den großen Vorteil der Unabhängigkeit eines Messfehlers von der Entfernung, aber auch die Herausforderung bei schnellen Signalen wie den elektromagnetischen Wellen, bei diesen Verfahren [381]. Je nach Art des Signals werden die Geräte als

- Lidar (*light detection and ranging*, Signal: gepulster Laserstrahl; z. B. Laserscanner),
- Radar (*radio detection and ranging*, Signal: Radio- oder Mikrowellen) oder
- Sonar (*sound navigation and ranging*, Signal: Schall oder Ultraschall; z. B. Echolot)

bezeichnet. In den Bereich der mittels Radio- oder Mikrowellen arbeitenden Geräte fällt auch die Ortung durch das globale Positionsbestimmungssystem (*global positioning system*, GPS), wobei bei diesem System kein Signal vom Messgerät ausgesandt wird, sondern die Zeitdifferenz kontinuierlich von Satelliten ausgesandter kodierter Radiosignale gemessen wird. Radar-Geräte sind auch zur Verfolgung von Personen nutzbar [382–384]. Zu den Lidar-Systemen zählen neben den Laserscannern auch die ohne bewegliche Teile auskommenden Laufzeit- bzw. TOF-Kameras (*time-of-flight camera*), die aber, statt wie beim Laserscanner den Abstand zeilen- oder rasterartig abzutasten, den Abstand einer ganzen Szene auf einmal meist mittels eines Photomischdetektors bzw. PMD-Sensors (*photonic mixing device*) mit einer aktuell maximalen Auflösung von ca. 200×200 Messeinheiten zur Laufzeitbestimmung erfassen [385].

2.3.9 Laserscanner

Die verbreitetste und schon lange eingesetzte Technik der auf Laufzeit basierenden Systeme zur Erkennung von Personen ist der Laserscanner. Das Abtasten bzw. Scannen kann dabei zwei- oder dreidimensional erfolgen. Im Bereich der Robotik wird typischerweise zweidimensional der Abstand horizontal bestimmt, um Kollisionen zu vermeiden. Bei Personenzählssystemen oder der Objektüberwachung werden vertikale Messungen zweidimensional an sogenannten Vorhängen durchgeführt. Wird bei einem dreidimensionalen Laserscanner zusätzlich die Intensität des reflektierten Strahls gemessen, so wird dieser abbildender Laserscanner genannt und erzeugt dabei ein Bild ähnlich einem Grauwertbild, so dass wie bei Stereokameras oder der Kinect mit Daten nur eines Systems neben der Tiefeninformation auch die Bildinformation vorliegt und zusätzlich zur Detektion von Personen herangezogen werden kann.

Der Vorteil eines Laserscanners gegenüber den optischen Techniken liegt in der kurzen Reaktionszeit, der relativen Unempfindlichkeit gegenüber der Umgebungsbeleuchtung und -textur, einer guten Tiefenauflösung insbesondere entfernter Objekte und der Abdeckung großer Bereiche, also einem großen Tiefenbereich für einzelne Systeme. Hingegen ist im sehr nahen Bereich bei der reinen Laufzeitmessung aufgrund der sehr kurzen zu erfassenden Zeitdifferenzen die Tiefenauflösung geringer als bei optischen Verfahren. Mittels eines Laserinterferometers können durch Messung der Phasenverschiebung jedoch auch sehr kleine Abstände bestimmt werden.

Auch bei den Verfahren basierend auf Daten eines Laserscanners wird für gewöhnlich zuerst der Hintergrund in den resultierenden Punktwolken eliminiert. Im Anschluss werden die restlichen Punkte des Vordergrundes örtlich gruppiert, um einzelne Objekte freizustellen. Die Gruppen werden dann auf mögliche Übereinstimmungen mit der Form einer Person im dreidimensionalen [386, 387] oder eines Körperteils wie den Beinen im zweidimensionalen [388–392] analysiert.

Einige Verfahren nutzen multimodale Ansätze unter Verwendung mehrerer Sensoren, meist Kameras, so z. B. [179, 182, 266, 393–396].

In der zur Durchführung von Laborexperimenten aufgebauten Plattform *pamela* (*pedestrian accessibility movement environment laboratory*) [158] (siehe Abb. 2.21) werden die Laufwege der Personen u. a. mittels eines Laserscanners detektiert [159].



Abb. 2.21: Laborplattform pamela (*pedestrian accessibility movement environment laboratory*) [158] zur Untersuchung von Bewegungen, Verhalten und Interaktion von Personen mit ihrer Umgebung; an der Decke und um die Plattform sind einige Sensoren zur Erfassung angebracht; Bild aus [158]

Ein System zur Personenzählung hatte ich bei einem Feldversuch zu Testzwecken im Einsatz. Dieses arbeitete mit zwei zweidimensionalen senkrechten Vorhängen und erfasste Personen ab einem Bodenabstand von 120 cm über ihr Höhenprofil. Über die Dopplung der Vorhänge konnte die Bewegungsrichtung ermittelt werden. Die Fehlerrate lag zwischen der einer ebenfalls getesteten monokularen Zählkamera und dem selbst entwickelten am besten abgeschnittenen Verfahren. Mehr zur Qualität der Ergebnisse und ein Vergleich zu den optischen Sensoren kann Abschnitt 3.6.7 entnommen werden. Weitere kommerzielle Anbieter von Personen erfassenden Laser-Systemen sind in Anhang E aufgeführt.

2.3.10 Weitere Systeme

Die verbreitetsten Verfahren zur genauen Ortung von Personen sind in den vorangegangenen Abschnitten besprochen worden und in der Tab. 2.3 gelb hinterlegt. Es gibt, wie auch in Tab. 2.3 zu sehen, eine Vielzahl von weiteren Systemen zur Erfassung von Personen [33], die ihre Anwendung in speziellen Feldern besitzen oder die bisher beschriebenen Methoden ergänzen können. Exakte Laufwege lassen sich noch mit Kontakt- [397, 398], Druck- [399] oder Vibrationssensoren [400], z. B. in Form von Kontaktplatten auf dem Boden, erfassen, andere geben nur grob Auskunft über den Aufenthaltsort bzw. Weg, den eine Person zurückgelegt hat.

Instrumentierte Techniken erlauben es, Personen über Kennungen der Geräte zu identifizieren und über größere Distanzen als den vorgenannten Methoden eindeutig zu verfolgen, dies jedoch meist mit geringerer Ortsauflösung. Ohne aktive Beteiligung der zu erfassenden Personen können nur die Personen verfolgt werden, die Geräte, insbesondere Mobiltelefone bzw. Smartphones, mit sich führen, so dass nur ein gewisser nicht notwendigerweise repräsentativer Prozentsatz an Personen detektiert werden kann. Hierzu können Bluetooth-, GSM- (*global system for mobile communications*) [401–405] oder

sensing modalities	signaling	presence	count	location	track	identity	ND
Uninstrumented							
Contact Sensors	passive	○	○	○	○		4
Pressure Sensors	passive	○	○	○	○	·	4
Chemosensors	passive	—	—	—	—	—	?
Photodetectors	passive	·	·	·	·		4
Cameras	either	○	○	○	○	○	0
Thermal Imagers	passive	○	○	○	○	·	0
Breakbeam Sensors	active	○	—	—	—		4
Scalar Range-Finders	active	○	—	—	—		0
Scanning Range-Finders	active	○	○	○	○	·	0
Tomographic Sensors	active	○	○	○	○	—	5
EF Sensors	active	○	○	○	○		4
Doppler-Shift Sensors	active	○	○	○	○	·	0
Motion Sensors	either	○	·	·	·		2
Seismic and Inertial Sensors	passive	○	·	·	·		3
Microphones	passive	○	·	·	·	·	1
Instrumented							
Wearable Inertial Sensors	passive	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	×
Wearable Environment Recognition	passive	Ⓢ	Ⓢ	Ⓢ	Ⓢ	Ⓢ	×
Wearable SS Device-to-Device Ranges	either	○	○	·	○	○	2
Wearable AA Device-to-Device Ranges	active	○	○	○	○	○	2
Wearable TOA/TDOA Dev.-to-Dev. Rang.	active	○	○	○	○	○	2
Wearable Doppler-Shift Sensors	active	○	○	○	○	○	2

○ = good performance ◦ = medium performance · = low performance
 — = plausible, but no detailed literature ? = no literature
 Ⓢ = requires communications (i.e. depends on the addition of a radio)
 × = not applicable: this is solely a self-sensing method, so no network is involved.

Tab. 2.3: Fähigkeiten verschiedener Sensorarten zur Wahrnehmung (*presence*), Zählung (*count*), Erfassung (*location*), Tracking (*track*) und individueller Erfassung (*identity*) von Personen; der ND-Wert gibt die Netzwerkdichte an, deren Wert etwas über die Anzahl an Sensoren pro erfasster Person aussagt; je kleiner der ND-Wert ist, desto besser; gelb eingefärbte Sensorarten wurden in diesem Kapitel ausführlich behandelt; rot eingefärbt sind die Spalten der Erkennung und des Verfolgens, die das unsere Experimente erfassende System mindestens ermöglichen muss; basiert auf Tabelle aus [33]

Wlan- (*wireless local area network*) [406–412] Signale der Mobiltelefone genutzt werden. Dabei ist man bei den GSM-Signalen für gewöhnlich auf die Unterstützung der Mobilfunkbetreiber angewiesen. Bluetooth-Scanner können hingegen einfach selber installiert werden [413–415].

Bei aktiver Beteiligung der Personen können GSM- oder Wlan-Signale auch zur Positionsbestimmung auf der Geräteseite genutzt werden. Eine verlässlichere Ortung ist bei Beteiligung der Personen jedoch mittels GPS-Signalen möglich, die aber nicht innerhalb von Gebäuden empfangbar sind [417–421].

Zur Verwendung der RFID- (*radio-frequency identification*) Technik zur Personendetektion sind diese ebenfalls mit RFID-Transpondern auszustatten. Die Transponder können bei passiver Auslegung günstig sein, erlauben aber keine genaue Ortung u. a. durch Signalstörungen umgebender Menschen [13, 422, 423]. Sie können aber aufgrund der Individualisierung gut zur Ergänzung anderer Verfahren herangezogen werden. Soll nur festgestellt werden, wer oder wie viele sich in größeren Bereichen, z. B. Räumen, aufhalten, ist die RFID-Technik eine günstige Technik [424, 425]. Auf ähnlicher Technik basierende kommerzielle Systeme zur Verfolgung von Personen oder anderen Objekten existieren [426, 427] und wurden im Rahmen der Vorbereitungen der eigenen Versuchsreihe Nr. 6 in Tab. 4.1 getestet und die Genauigkeit als für unsere Zwecke ungenügend bewertet.

Systeme, die zur Erfassung der relativen Bewegung im Raum und damit der relativen Position referenzlos ohne Nutzung von externen Messsystemen oder Ortungssignalen auskommen, werden Trägheitsnavigationssysteme bzw. inertielle Navigationssysteme (*inertial navigation system*, INS) genannt. Hierzu müssen alle sechs kinematischen Freiheitsgrade (drei translatorische und drei rotatorische) eines Objektes erfasst werden. Dies wird mittels Inertialsensoren (z. B. Gyroskop, Beschleunigungssensor, Neigungsmesser) kombiniert in einer inertialen Messeinheit (*inertial measurement unit*, IMU) durchgeführt, die anhand von intern verbauten, quantitativ bekannten Massen aufgrund der Massenträgheit die Beschleunigung und Drehrate ermitteln kann. Sie ermöglicht ohne örtliche Begrenzung, in sowie außerhalb eines Gebäudes und ohne externe Hilfsmittel in Echtzeit die Bewegung einer Person festzuhalten [412, 428–432].

Die Stärken liegen darin, dass keine Abschirmung und Beeinflussung möglich, keine Kommunikation nötig, der Arbeitsbereich unbegrenzt, und die Bauform sehr klein realisierbar ist. Der große Nachteil liegt in der kumulativen Auswirkung von Messfehlern (lineare Zunahme durch Nutzung von Kompensationsstrategien) (siehe Abb. 2.22). Daher werden diese Systeme gerne mit fest im Weltkoordinatensystem verorteten Systemen [433] wie z. B. dem globalen Positionsbestimmungssystem GPS gekoppelt [434], die zwar nicht die Genauigkeit besitzen, aber eine absolute Positionsbestimmung erlauben. Daher spricht man auch von Koppelnavigation (*pedestrian dead reckoning*, PDR) [435]. Soll die Navigation innerhalb von Gebäuden durchgeführt werden, können die RFID-Technik [436] oder Wlan-Signale [437–439] ergänzend genutzt werden.

Firmen, die Produkte für die inertielle Navigation anbieten, sind z. B. InterSense mit dem NavShoe [440] und die Firma Xsens mit MTw [441].

Wie bei den inertialen Navigationssystemen ist auch eine Kopplung der in den vorherigen Abschnitten beschriebenen Techniken mit den instrumentierten Systemen denkbar und kann die Vorteile beider Systeme verbinden. So wäre eine individuelle Erfassung von detaillierten Laufwegen mittels einer gemeinsamen Nutzung von RFID- und Kamera-Technik [443, 444] denkbar oder eine großräumige individuelle Verfolgung mit zumindest teilweise exakten Laufwegen durch Kombination von GPS oder Bluetooth und Kamera-Techniken realisierbar. Die in [13, 154] beschriebenen Experimente zur Gebäuderäumung nutzten die RFID-Technik zur Erfassung von Raum- oder Ebenenwechsel im Treppenhaus (siehe Abb. 2.23). Parallel wurden zur Personendetektion Kameras eingesetzt.

2.3.11 Vergleich

Die vorgestellten Systeme haben, wie im Laufe dieses Kapitels beschrieben, unterschiedliche Stärken und sind daher in verschiedenen Bereichen bei der Erkennung von Personen im Einsatz.

Bei der für uns wichtigen Erfassung von exakten Laufwegen in dichten Menschenmassen ist bei aus-

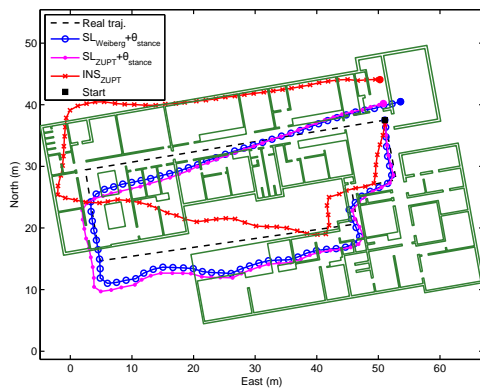


Abb. 2.22: Laufweg durch ein Gebäude bestimmt mittels einer inertialen Messeinheit und unterschiedlichen Strategien zur Korrektur; diese nutzen u. a. die Annahme einer festen Schrittlänge; Bild aus [433]



Abb. 2.23: (links) gelb umkreiste RFID-Antenne und Kamera im Treppenhaus zur Erfassung von Ebenenwechsel während einer Evakuierung; (rechts) mit passiven RFID-Transpondern ausgestattete rote Ausweise der Versuchsteilnehmer; Bilder aus [154]

Eigenschaften	Kamera	Stereokamera	Mehrkamerasystem	Infrarotkamera	Wärmebildkamera	RGB-D-Sensor	Laufzeit-Kamera	Laserscanner
Personendetektion	+	++	++	+	+	++	+	+
Personenverfolgung	++	++	++	+	+	++	o	o
Personenidentifikation	+	+	++	o	o	+	-	-
Auflösung	++	++	++	+	o	+	-	++
Tiefenauflösung	-	+	o	-	-	+	+	++
Reichweite	+	+	++	+	+	o	o	++
Messwiederholrate	+	o	o	+	o	+	+	o
Umgebungsanforderung	-	-	-	o	o	+	++	++
Kosten	++	+	+	+	-	++	o	-
Größe / Gewicht	++	+	+	+	o	+	+	-

Tab. 2.4: Qualitative Bewertung der in diesem Kapitel vorgestellten Techniken; Tiefenauflösung ist die Genauigkeit in der Tiefenbestimmung (hier bezogen auf den Bereich zwischen 3 m und 15 m); Messwiederholrate schließt sie Bestimmung der 3D-Information mit ein; Umgebungsanforderungen umfasst u. a. Beleuchtung oder Textur (++ sehr gut, + gut, o mäßig, - schlecht)

reichender Beleuchtung dieameratechnik (einschließlich Stereokameras) am weitesten verbreitet und am verlässlichsten. Teixeira schreibt in [33]: „*For uninstrumented scenarios, the best modality across the board is vision (i.e. cameras and other imagers). Computer vision is far ahead from other instrumented modalities not only with respect to spatial-resolution and precision metrics, but also in terms of having the most field-tested solutions.*“

Bei der eindeutigen Identifizierung von Personen können instrumentierte Systeme dieameratechnik sinnvoll ergänzen (siehe Tab. 2.3), wobei mit dem in dieser Arbeit gezeigten Codemarker (siehe Abb. 3.9 5) ebenfalls eine Identifizierung möglich ist.

Ein quantitativer Vergleich der unterschiedlichen Systeme ist nicht durchführbar, da die Resultate der einzelnen Techniken und sie einsetzende Verfahren stark von den Randbedingungen abhängen.

Dieameratechnik ist günstig, weit verbreitet und besitzt eine hohe Auflösung, Infrarot- und Wärmebildkameras können hingegen bei schlechten Lichtverhältnissen Personen identifizieren. Bei den Verfahren, die auf monokularerameratechnik basieren, ist bei großen Personendichten die Fehlerrate hoch, kann aber durch die Sicht von oben und den Einsatz von Stereokameras oder Mehrkamerasystemen wesentlich verbessert werden, da die Form als Merkmal zur Detektion herangezogen werden kann.

Weitere in diesem Kapitel angesprochene Techniken zur Erfassung von 3D-Informationen sind RGB-D-Sensoren, wie die Kinect, TOF-Kameras und Laserscanner. Auch hier zeigt sich, dass jede Technik ihre Stärken hat (vgl. Tab. 2.4). Der Laserscanner kann auch entfernte Personen noch gut dreidimensional abbilden, kann einen großen Bereich erfassen und ist auf keine Oberflächenstruktur angewiesen, und mit der Kinect steht ein sehr kostengünstiges System zur Verfügung, das auch im Dunkeln gleich gute Ergebnisse liefert und somit z. B. Beobachtungen bei Großveranstaltungen, wie Konzerten, mit geringer Beleuchtung ermöglicht.

Eine qualitative Bewertung der vorgestellten Systeme sind zusammengefasst der Tab. 2.4 zu entnehmen, in die für die 3D-Kameras die Bewertungen aus [445] eingeflossen sind.

2.4 Tracking

Wie für die in diesem Abschnitt vorgestellten Erkennungsverfahren könnte ebenfalls eine Übersicht über Verfahren zur Verfolgung von Personen folgen. Da die die Versuche erfassende Technik so ausgewählt und positioniert wurde, dass aufgrund von geringer Verdeckung und hohem Kontrast im Bereich des zu verfolgenden Kopfes bereits mit einem fertig in OpenCV [446, 447] implementierten Verfahren, dem pyramidenförmigen iterativen Lucas-Kanade-Merkmals-Verfolger (*pyramidal iterative Lucas Kanade feature tracker*) [448] basierend auf dem Kanade-Lucas-Tomasi-Merkmals-Verfolger (*Kanade Lucas Tomasi feature tracker*, KLT) [449–451], sehr gute Ergebnisse erzielt wurden, soll hier nur auf einige wichtige Arbeiten und Übersichtsartikel in diesem Gebiet verwiesen werden: [68, 452–455]. Manche oben aufgeführte Erkennungsverfahren nutzen die zeitliche Veränderung der Erscheinung eines Objektes zur Entscheidung, ob es sich um einen Menschen handelt, und brauchen somit bereits für die Erkennung eine Verfolgung der Person.

3

Erfassung von Laufwegen

In diesem Abschnitt sollen die selbst entwickelten Verfahren zur Erfassung von Laufwegen von Personen im Detail vorgestellt werden. Alle Verfahren nutzen ausschließlich Kameras mit Sensoren, die elektromagnetische Wellen des für den Menschen sichtbaren Lichtes erfassen. Somit können die Verfahren zur Erkennung der Personen von Erfahrungen der menschlichen Wahrnehmung abgeleitet werden.

Für die Ermittlung der Trajektorien sind mehrere aufeinanderfolgende Schritte nötig. Diese gemäß Abb. 3.2 benannten Schritte sind:

- a) Ausgehend von einem Bild einer aufgenommenen Bildfolge
- b) muss zuerst eine Kalibrierung durchgeführt werden, um die optische Verzerrung des Linsensystems der Kamera zu verringern (siehe Abschnitt 3.1.2) und
- c) um die Kamera in ein Weltkoordinatensystem einzubetten, so dass eine metrische Position eines Objektes an einem Bildpunkt bestimmt werden kann (siehe Abschnitt 3.1.1).
- d) Die Lokalisierung der Personen wurde bei den bisherigen Versuchsreihen aufgrund der Robustheit und Genauigkeit mittels Marker durchgeführt (siehe Abschnitt 3.2). Für Feldstudien wird in Abschnitt 3.5 auch ein Verfahren vorgestellt, das ein mittels einer Stereokamera bestimmtes Disparitätsbild unmittelbar zum Auffinden von Formen eines Menschen nutzt.
- e) Nach der Detektion der Personen müssen diese über die Bildsequenz hinweg verfolgt werden (siehe Abschnitt 3.3).

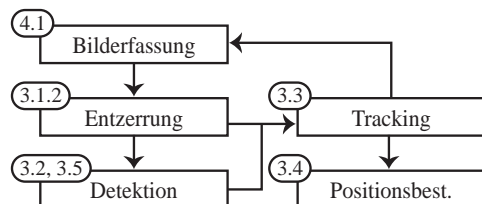


Abb. 3.1: Schema des Ablaufs der Erfassung der Laufwege vom Bild über die Entzerrung, einer möglichen Detektion, dem Tracking und der Positionsbestimmung der Person in Weltkoordinaten; die Detektion kann übersprungen werden, wenn nicht in jedem Bild eine Detektion durchgeführt wird oder eine Person in einem Bild nicht erkannt wird; die Zahlen geben den Abschnitt an, in denen der entsprechende Punkt erläutert wird

- f) Für die Positionsbestimmung (siehe Abschnitt 3.4) wird aufgrund der perspektivischen Verzerrung der Kameraoptik die Entfernung des zu detektierenden Objektes benötigt. Die Entfernung kann im einfachsten und ungenauen Fall aus dem Abstand der senkrecht zur Bewegungsebene aufgehängenen Kamera und dem Kopf in Höhe der durchschnittlichen Personengröße bestimmt werden. Um den aus dieser Annahme gleich großer Personen resultierenden Fehler zu verringern, werden zum Teil spezielle Marker (siehe Abschnitt 3.4.1) oder Stereokameras (siehe Abschnitt 3.4.2) eingesetzt. Die Stereokameras ermöglichen zudem die Positionsbestimmung in Personenströmen auf geneigten Grundflächen, z. B. an Treppen.

Die Bildfolgen in Abb. 3.2 und 3.3 zeigen diese Schritte zum einen für ebene Versuche mit farbkodierter Angabe der Personengröße (siehe Video Nr. 50) und zum anderen für Versuche in einer Tribüne unter Ausnutzung des Disparitätsbildes einer Stereokamera (siehe Video Nr. 44).

Die Güte der ermittelten Laufwege wird in Abschnitt 3.6 analysiert. Alle hier vorgestellten Verfahren sind implementiert worden und über die Software `PeTrack` [456] frei zugänglich.

3.1 Kalibrierung

Bevor Daten aus Bildern gewonnen werden können, muss der Abbildungsprozess kalibriert werden. Zum einen, um die geometrische Verzerrung bzw. Verzeichnung des optischen Systems, die für eine Vermessung wichtigste Aberration, zu beheben, und zum anderen, um das Bildkoordinatensystem in das Weltkoordinatensystem einzubetten, so dass eine Längenmessung im Bild vorgenommen werden kann.

Die Verzerrungen sind insbesondere bei den oftmals verwendeten Weitwinkelobjektiven vor allem zum Rand hin besonders groß (siehe Abb. 3.7 a, 3.2 a und 3.3 a und b). Die Weitwinkelobjektive sind trotzdem verwandt worden, um einen möglichst großen Bereich einsehen zu können (siehe Abschnitt 4.1.1.2 und 4.1.1.3).

3.1.1 Vermessung

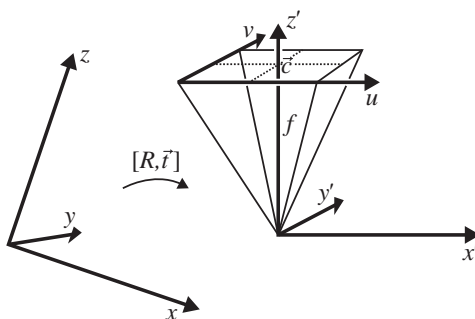


Abb. 3.4: Koordinatentransformation zwischen Welt-, Kamera- und Bildkoordinatensystem

Zur Beschreibung der Abbildung der realen Welt auf die Bildebene, in der der Sensor einer Kamera liegt, kann das Idealmodell einer Lochkamera herangezogen werden. Dieses Modell entspricht vom Resultat her der perspektivischen Abbildung auf die Schärfeebene im Schärfentiefebereich einer Linsenkamera. Die inverse Abbildung ermöglicht die Gewinnung von metrischen Informationen aus dem Bild.

Sei $\vec{x} \in \mathbb{R}^4$ die homogene Erweiterung eines Punktes $\vec{x} = (x \ y \ z)^T$ im Weltkoordinatensystem und $\vec{u} \in \mathbb{R}^3$ die homogene Erweiterung des Bildpunktes bzw. Pixels $[px] \ \vec{u} = (u \ v)^T$, so werden folgende Daten für die Abbildung des realen

Punktes \vec{x} auf den Bildpunkt \vec{u} benötigt (siehe Abb. 3.4):

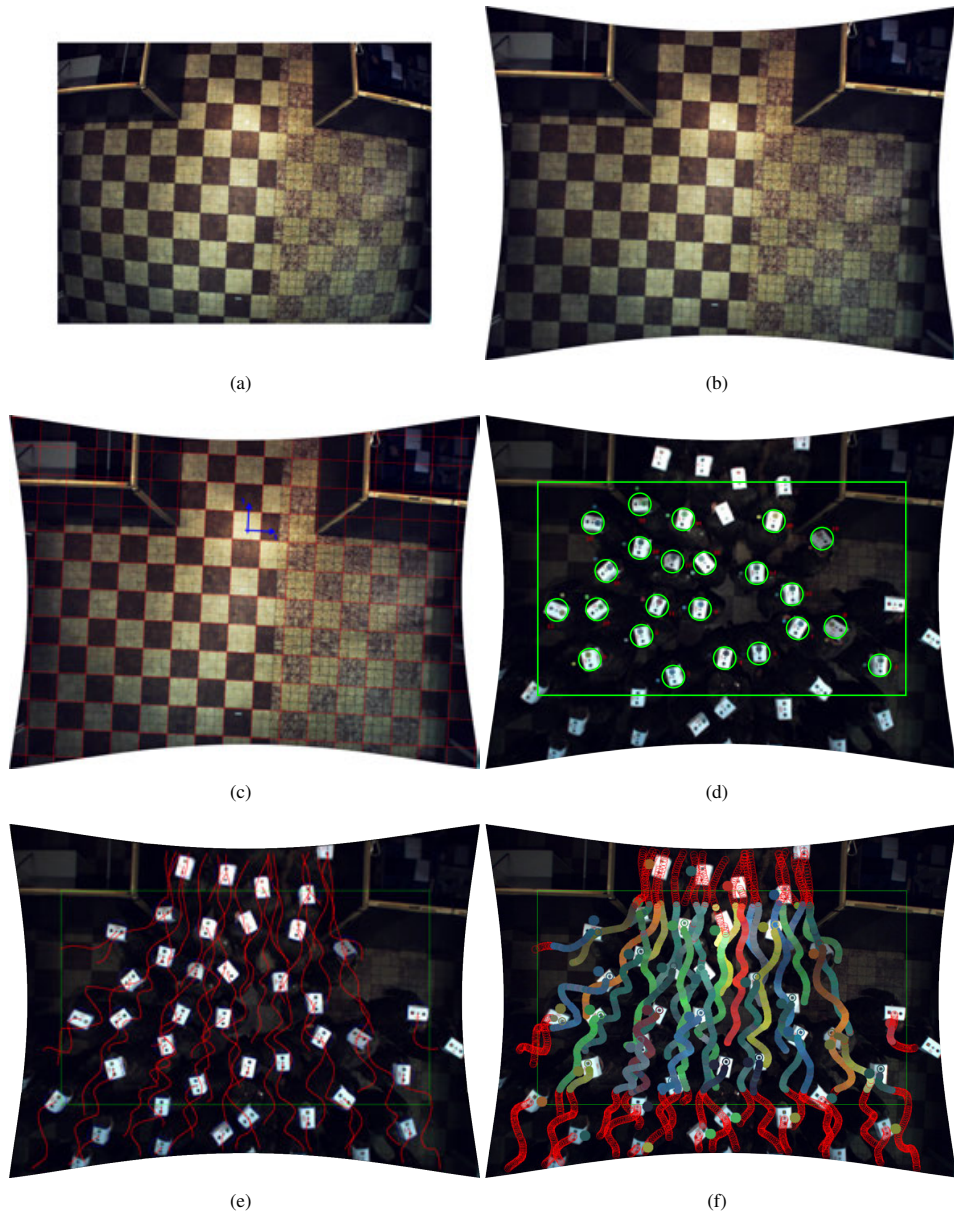


Abb. 3.2: Schritte zur Erfassung von Laufwegen, hier von einem Kasernen-Versuch (Versuchsreihe Nr. 3) mit Kasernen-Marker (Nr. 1) (siehe Video Nr. 50); alle Bilder, in denen keine Personen mit Markern zu sehen sind, wurden zur besseren Darstellung aufgehellt; (a) zeigt das aufgenommene Bild ohne Versuchspersonen wie es von der Kamera Nr. 1 geliefert wurde; (b) das entzerrte Bild (siehe Abschnitt 3.1.2); (c) mit überlagertem Gitter zur visuellen Qualitätskontrolle und dem eingetragenen Koordinatensystem zur Bildmessung (siehe Abschnitt 3.1.1); (d) zeigt grün umkreist die über die Marker im grünen Rechteck erkannten Personen, für die zudem eine fortlaufende Nummer, das Farbmarkerelement und die Durchschnittsfarbe angezeigt wird (siehe Abschnitt 3.2.1); (e) zeigt als roten Pfad den Laufweg der letzten Sekunden jeder erkannten Person (siehe Abschnitt 3.3); (f) zeigt entlang des Pfades die im jeweiligen Bild erfasste Farbe zur Berechnung der Durchschnittsfarbe für die Positionsbestimmung im Weltkoordinatensystem (siehe Abschnitt 3.4.1)

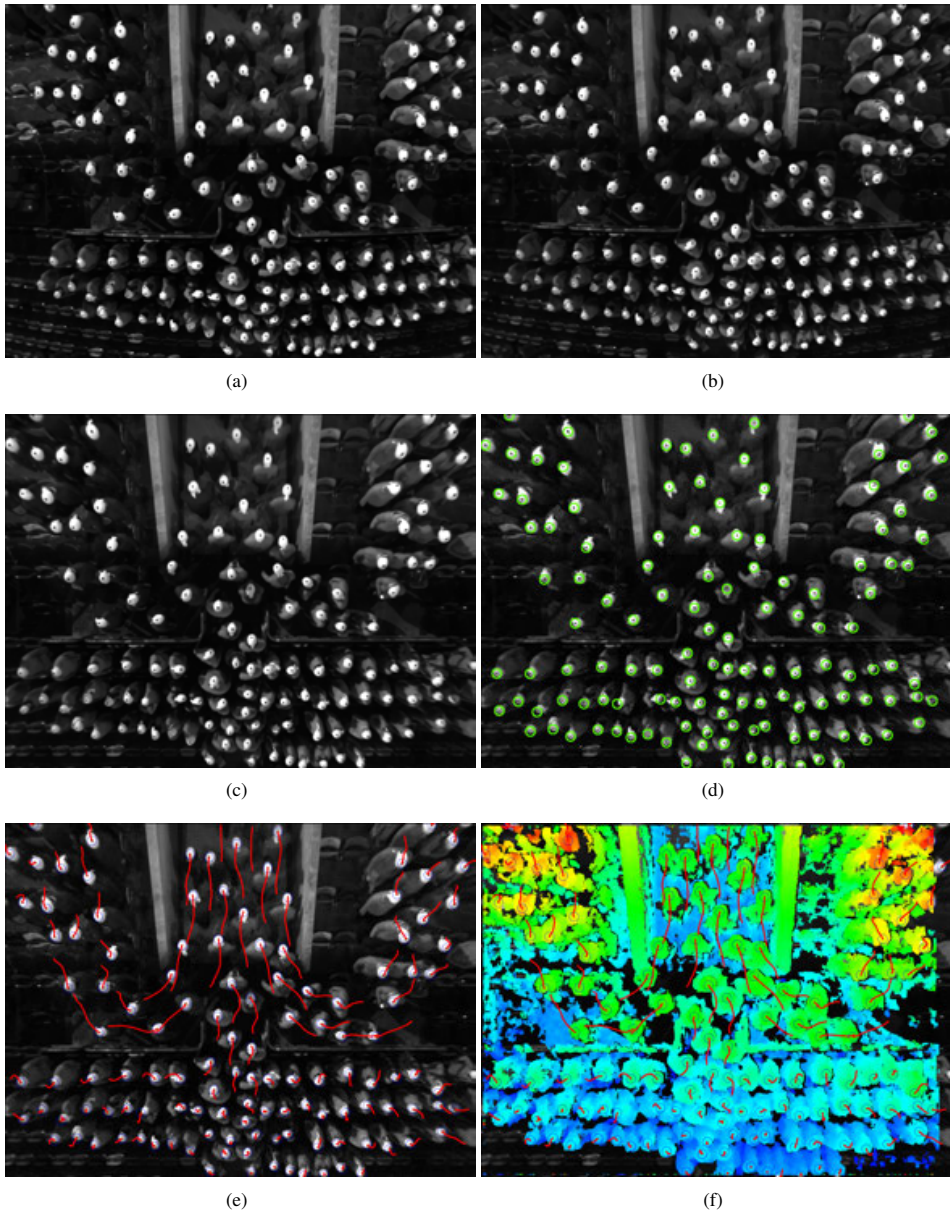


Abb. 3.3: Schritte zur Erfassung von 3D-Laufwegen, hier von einem Hermes-Versuch (Versuchsreihe Nr. 6) mit Hermes-Markern (Nr. 2) beim Verlassen einer Tribüne durch ein Mundloch; (a, b) zeigen beide versetzten Blickwinkel der eingesetzten Stereokamera Nr. 4 (siehe Tab. 4.2); (c) ist das entzerrte linke Bild a (siehe Abschnitt 3.1.2); in (d) wurde die Detektion mittels Marker durchgeführt, wobei der Mittelpunkt der grünen Kreise die Pixelposition jeder Person anzeigt (siehe Abschnitt 3.2); (e) zeigt als roten Pfad den Weg jeder Person in der letzten Sekunde (siehe Abschnitt 3.3); (f) besitzt zudem als Überlagerung die farbkodierte Disparität der versetzten Blickwinkel und damit das Höhenfeld bzw. die Entfernung zur Kamera (rot \cong 5 m bis blau \cong 10 m) für die Positionsbestimmung im Weltkoordinatensystem (siehe Abschnitt 3.4.2)

- extrinsische Parameter zur Beschreibung der Abbildung vom Welt- in das Kamerakoordinatensystem (Orientierung der Kamera bzw. äußere Orientierung):
 - Rotation $R \in \mathbb{R}^{3 \times 3}$ (Aus- bzw. Blickrichtung),
 - Translation $\vec{t} \in \mathbb{R}^3$ [m] (Lage);
- intrinsische Parameter zur Beschreibung der Abbildung vom Kamera- in das Bildkoordinatensystem (kamerainterne bzw. kameraabhängige Parameter bzw. Parameter der inneren Orientierung):
 - Brennweite (*focal length*) $f \in \mathbb{R}$ [m] (Abstand der Bildebene zum Ursprung des Kamerakoordinatensystems),
 - Skalierungsfaktor $s_x, s_y \in \mathbb{R}^{>0}$ [px/m] (Faktor zur Umrechnung von Pixelwerten in metrische Maße),
 - Kamerahauptpunkt (*principal point*) $\vec{c} = (c_x, c_y)^T \in \mathbb{R}^2$ [px] (Zentrum der Verzerrung und Schnittpunkt der optischen Achse mit der Bildebene),
 - radiale Verzerrungsparameter $k_1, k_2 \in \mathbb{R}$,
 - tangentielle Verzerrungsparameter $p_1, p_2 \in \mathbb{R}$.

Dabei liegen alle Werte im Reellen und nicht in diskreten Pixelkoordinaten vor, da alle anschließenden Verfahren auf Subpixel-Ebene arbeiten. Alle in der weiteren Arbeit verwendeten Symbole wie u. a. die intrinsischen und extrinsischen Parameter sind in Anhang F.1 zusammengefasst aufgeführt.

Die Abbildung eines realen Punktes \vec{x} im Weltkoordinatensystem auf einen Bildpunkt \vec{u} lautet damit [457, 458]:

$$\vec{u} = A[R, \vec{t}]\vec{x} \quad \text{mit} \quad A = \begin{pmatrix} s_x f & 0 & c_x \\ 0 & s_y f & c_y \\ 0 & 0 & 1 \end{pmatrix}. \quad (3.1)$$

Das Weltkoordinatensystem ist einheitenlos bzw. die Einheit ist frei wählbar, wohingegen das Bildkoordinatensystem die sensorspezifische Einheit Pixel hat. Die Einheit des Weltkoordinatensystems wird hier über Skalierungsfaktoren s_x, s_y festgelegt, die das Pixel-Meter-Verhältnis in x - und y -Richtung setzen. Dabei ist $s_x \neq s_y$ für nicht quadratische Pixel, beispielsweise für Kameras mit einem Pixelseitenverhältnis (*pixel aspect ration*) ungleich 1:1 in Tab. 4.2.

$A \in \mathbb{R}^{3 \times 3}$ wird Kalibrierungs- oder auch intrinsische Matrix genannt und beschreibt die Abbildung vom Kamera- in das Bildkoordinatensystem:

$$\begin{array}{ccccc} \text{Weltkoordinate} & & \text{Kamerakoordinate} & & \text{Bildkoordinate} \\ \vec{x} & \xrightarrow{\quad [R, \vec{t}] \quad} & \vec{x}' & \xrightarrow{\quad A \quad} & \vec{u} \\ & \text{äußere Orientierung} & & \text{innere Orientierung} & \end{array}$$

Bei der Bildanalyse sind wir an dem umgekehrten Prozess interessiert. Für einen Bildpunkt ergibt sich bei vorheriger Entzerrung der reale Punkt

$$\vec{x} = R^{-1}(\vec{x}' - \vec{t}) \quad \text{mit} \quad \vec{x}' = A^{-1}\vec{u}. \quad (3.2)$$

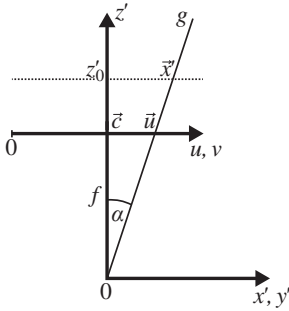


Abb. 3.5: Lagebestimmung eines Bildpunktes \bar{u} im Kamerakoordinatensystem in der Ebene $E : z' = z'_0$

\bar{x}' ist der in das Kamerakoordinatensystem überführte reale Punkt \bar{x} bzw. in der inversen Betrachtung der überführte Bildpunkt \bar{u} :

$$\bar{x}' = \begin{pmatrix} \frac{1}{s_x f} & 0 & -\frac{c_x}{s_x f} \\ 0 & \frac{1}{s_y f} & -\frac{c_y}{s_y f} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \bar{u} \\ \bar{v} \\ \bar{w} \end{pmatrix}. \quad (3.3)$$

Diese inverse projektive Abbildung ist aufgrund der homogenen Erweiterung \bar{w} unterbestimmt und hat im Allgemeinen als Lösung eine Gerade g vom Projektionszentrum durch den Bildpunkt.

Für eine eindeutige Lösung bedarf es einer weiteren Einschränkung, z. B. auf eine Ebene, in der der gesuchte reale Punkt liegt. Geschieht die Messung wie in dieser Arbeit parallel zur Bildebene und kann damit die einschränkende Ebene durch $E : z' = z'_0$ beschrieben werden, so ergibt sich nach Normierung der homogenen Erweiterung aus Gleichung 3.3:

$$\bar{x}' = \frac{z'_0}{f} \begin{pmatrix} \frac{1}{s_x} (u - c_x) \\ \frac{1}{s_y} (v - c_y) \\ 1 \end{pmatrix}. \quad (3.4)$$

In Abb. 3.5 ist dies nach Verschiebung des Kamerahauptpunktes in den Ursprung und unter Vernachlässigung der Skalierungsfaktoren komponentenweise durch den Strahlensatz nachzuvollziehen.

Sei $f' = s_x f$ die in Pixel umgerechnete Brennweite, so gilt für den Winkel α zwischen der optischen Achse und der Geraden g :

$$\tan \alpha = \frac{z'_0}{f} = \frac{u - c_x}{f'}.$$

Gleichung 3.4 lässt sich somit für den Kreis in der Ebene E in einem Winkelabstand $\alpha = \arctan \frac{u - c_x}{f'}$ zwischen der optischen Achse und dem Kreis mit $\beta \in [0, 2\pi[$ auch schreiben als:

$$\bar{x}' = z'_0 \tan \alpha \begin{pmatrix} \frac{1}{s_x} \sin \beta \\ \frac{1}{s_y} \cos \beta \\ 1 \end{pmatrix}. \quad (3.5)$$

Für die Anwendung in dieser Arbeit zur Positionsbestimmung von Personen überkopf ist aufgrund der orthogonalen Sicht zum Festlegen der extrinsischen Parameter R und \bar{t} lediglich ein 2D-Koordinatensystem auf dem Boden zu platzieren.

Das Platzieren des Koordinatensystems und das Festlegen des Skalierungsfaktors geschieht in der entwickelten Software `PeTrack` manuell (siehe Anhang D.2 und Abb. 3.2 c). Dabei wird statt eines koordinatenabhängigen Skalierungsfaktors eine koordinatenabhängige Brennweite f_x, f_y für die Verarbeitung von nicht quadratischen Pixel verwandt. Dass dies keine Einschränkung bedeutet, ist Gleichung 3.1 zu entnehmen ($s_x f = s f_x, s_y f = s f_y$).

Der Abstand $a \in \mathbb{R}$ [m] der Kamera vom Boden kann bei vorheriger Ausmessung in `PeTrack` explizit angegeben werden, ist aber auch aufgrund des auf dem Boden platzierten Koordinatensystems, der

Angabe der Brennweite f' [px] und dem Skalierungsfaktor s [m/px] berechenbar. Sei $d_k \in \mathbb{R}$ [px] die Anzahl der Pixel des manuell platzierten Koordinatensystems, so ergibt sich gemäß Strahlensatz die Aufhängehöhe von

$$\frac{a}{sd_k} = \frac{f'}{d_k} \Leftrightarrow a = sf'. \quad (3.6)$$

Der Kamerahauptpunkt \vec{c} liegt in der Bildmitte, wenn die Optik genau mittig und senkrecht zum Bildsensor verbaut wurde und damit die optische Achse durch die Bildmitte verläuft.

Bei dem Bildkoordinatensystem kann es je nach Implementierung nötig sein, von einem rechts- auf ein linkshändiges Koordinatensystem zu wechseln, wenn die Pixelkoordinaten positiv sind und ihren Ursprung in der oberen linken Ecke haben. Dies ist durch eine Spiegelung der v -Achse an der Bildmitte realisierbar.

3.1.2 Entzerrung

Wie eingangs des Abschnittes 3.1 beschrieben, muss für die Kalibrierung ebenfalls die Verzerrung des Bildes durch das optische System beachtet werden. Um die Linsenverzerrung zu erfassen, wird das im vorherigen Abschnitt beschriebene Modell der Lochkamera um radiale und tangential Verzerrungsparameter bis zur Ordnung vier erweitert [341, 459].

Die Lage des Bildpunktes \vec{u} wird dabei vor der Vermessung abhängig vom Abstand $r \in \mathbb{R}^{>0}$ zum Kamerahauptpunkt \vec{c} wie folgt abgebildet:

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} u(1 + k_1 r^2 + k_2 r^4) + 2p_1 uv + p_2(r^2 + 2u^2) \\ v(1 + k_1 r^2 + k_2 r^4) + p_1(r^2 + 2v^2) + 2p_2 uv \end{pmatrix}. \quad (3.7)$$

k_1, k_2 und p_1, p_2 sind die in Abschnitt 3.1.1 aufgeführten intrinsischen radialen und tangentialen Verzerrungsparameter.

Zur Illustration der Wirkung der einzelnen Verzerrungsparameter siehe Abb. 3.6.

Die selbst entwickelte Software erlaubt vor der Entzerrung das Hinzufügen eines Bildrandes und damit eine Vergrößerung des Bildes, um bei positiven radialen oder tangentialen Parametern ungleich Null keine Bildinformationen zu verlieren (siehe Abb. 3.2 b).

Die Entzerrung beruht auf einem eingeschränkten Modell, so dass auch bei optimaler Bestimmung der Parameter nicht alle optischen Effekte abgebildet werden können. Abb. 3.2 c zeigt ein rotes kartesisches Gitter, das über die auf dem Boden befindlichen Fliesen gelegt wurde und zum Rand bzw. zu den Ecken eine steigende Abweichung zeigt. In der späteren Fehlerbetrachtung in Abschnitt 3.6.1 wird hierauf näher eingegangen.

Die Festlegung der Verzerrungsparameter kann in der Software *PeTrack* manuell geschehen, es wird jedoch auch die automatische Bestimmung der intrinsischen Parameter ermöglicht. Das hierzu eingesetzte Verfahren [460] ist wie folgt aufgebaut:

1. Aufnahmen eines bekannten ebenen Musters (*pattern*) mit der zu kalibrierenden Kamera,
2. Extraktion der Merkmalspunkte (*feature points*) des Musters in jeder Aufnahme,
3. Minimierung des Reprojektionsfehlers zwischen parametrisiertem Modell und der Realität (Merkmalspunkte in der Aufnahme).

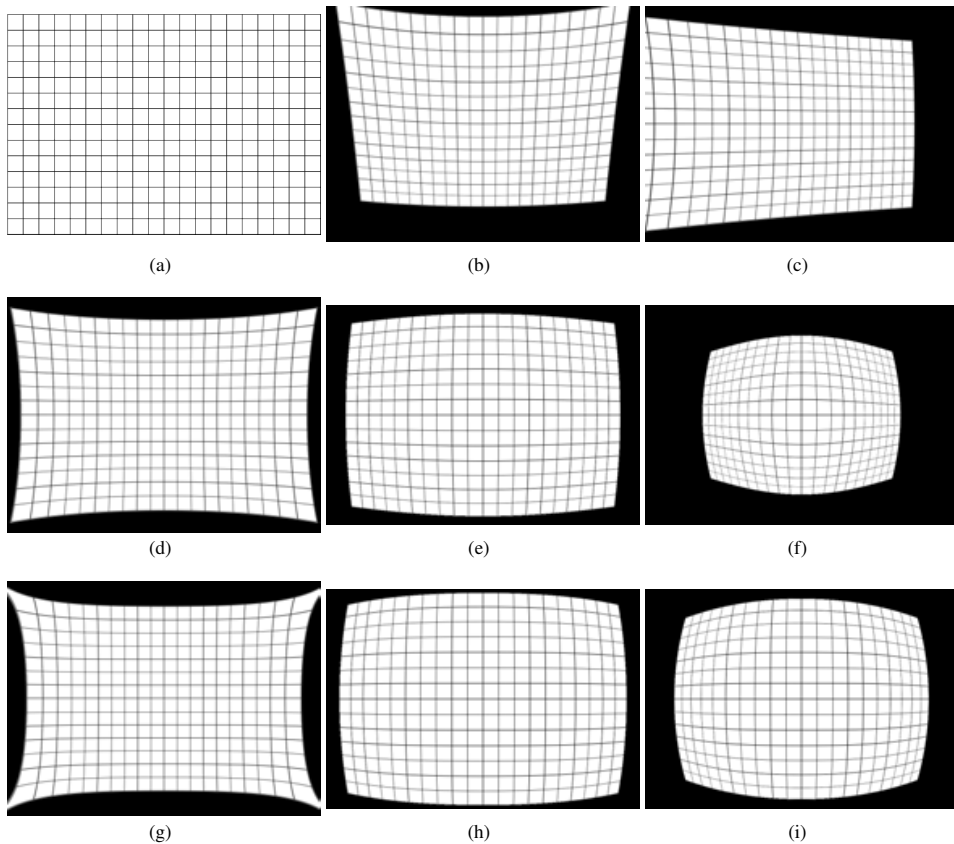


Abb. 3.6: Wirkung der Verzerrungsparameter in Gleichung 3.7: alle nicht beteiligten Parameter sind Null; (a) Bild eines kartesischen Gitters ohne Verzerrung; (b) tangentialer Parameter $p_1 > 0$; (c) $p_2 > 0$; (d) radialer Parameter der Ordnung zwei $k_1 < 0$; (e, f) $k_1 > 0$; (g) radialer Parameter der Ordnung vier $k_2 < 0$; (h, i) $k_2 > 0$

Bei der Aufnahme des Musters muss darauf geachtet werden, dass sich die Kamera im selben Zustand befindet wie bei den auszuwertenden Aufnahmen. So muss u. a. die gleiche Brennweite (Zoom) eingestellt, aber auch das Objektiv identisch auf der Kamera verschraubt sein. Am praktikabelsten ist dies, wenn die Kalibrierung direkt bei der Erstellung der auszuwertenden Aufnahmen durchgeführt wird.

Das wellenlose Aufbringen des auf Papier ausgedruckten als planar angenommenen Musters auf eine ebene Fläche ist in der Praxis schwer umzusetzen (leichte Wellen sind auch in Abb. 3.7 zu erkennen). Die besten Ergebnisse wurden erzielt, indem das Kalibrierungsmuster zwischen zwei Glasscheiben fixiert wurde.

Die Merkmalspunkte aller Aufnahmen zusammen sollten die Bildfläche gleichmäßig abdecken, denn nur die Bildbereiche, in denen Merkmalspunkte liegen, können bei der Bestimmung der Verzerrungsparameter berücksichtigt werden. Mit zwei Aufnahmen ist eine Bestimmung möglich und bereits mit

drei Aufnahmen eine gute Abschätzung der Parameter erzielbar [461].

Seien $g_{i,j}$ die extrahierten m Merkmalspunkte aus n Aufnahmen und die Funktion $g_{R,\vec{x},A,k_1,k_2,p_1,p_2}(\vec{x}) = \vec{u}$ die Abbildung der realen Koordinate auf die Bildkoordinate aus Gleichung 3.1 vor und nach der Überführung der homogenen in die inhomogene Koordinate, so lässt sich damit die Minimierung des Reprojektionsfehlers darstellen als

$$\operatorname{argmin} \sum_{i=1}^n \sum_{j=1}^m \|g_{i,j} - g_{R,\vec{x},A,k_1,k_2,p_1,p_2}(\vec{x})\|. \quad (3.8)$$

Damit wird der Abstand der Punktkorrespondenzen des Abbildes des bekannten Musters zwischen der optimalen Lage gemäß Modell und der Lage in den Aufnahmen minimiert. Für diese Optimierung wird wie in [460] die Maximum-Likelihood-Schätzung angewendet.

Vor der automatischen Bestimmung können bestimmte Eigenschaften der Kamera in `PeTrack` vorgegeben und bei der Optimierung berücksichtigt werden, um Vorwissen über die Kamera einfließen lassen zu können. So kann für die Parameter festgelegt werden, dass es sich um quadratische Pixel handelt, also $f_x = f_y$ ist, der Kamerahauptpunkt in der Bildmitte liegt, also die Optik nicht im Verhältnis zum Bildsensor verschoben ist, und keine tangentiale Verzerrung vorliegt, also die optische Achse senkrecht auf dem Bildsensor steht und damit $p_1 = p_2 = 0$ ist.

Durch die Nutzung eines bekannten Musters ist die relative Lage der Merkmalspunkte exakt bekannt und erlaubt neben der Bestimmung der inneren Orientierung auch die Bestimmung der äußeren Orientierung der Kamera im Verhältnis zum Muster in jeder Aufnahme bis auf den Skalierungsfaktor.

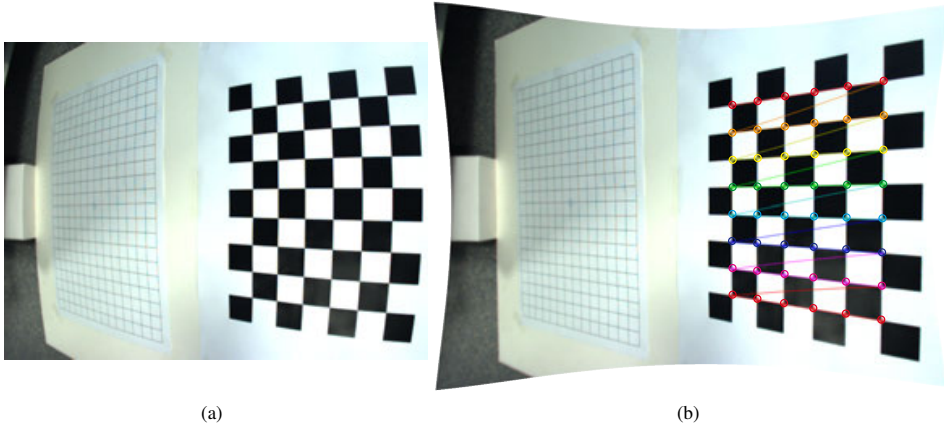


Abb. 3.7: Aufnahme des verwendeten schachbrettartigen Musters mit den Kreuzungspunkten als Merkmalspunkte; (b) detektierte Merkmalspunkte sind zeilenweise unterschiedlich farbig gekennzeichnet und die Entzerrung mit den ermittelten Verzerrungsparametern durchgeführt worden

Abb. 3.7 zeigt eine Aufnahme des verwendeten Musters. Die Merkmalspunkte sind die Kreuzungen des Schachbrettmusters, deren genaue Erfassung unempfindlich gegenüber Helligkeitsänderungen ist. In Abb. 3.7 b sind die detektierten Merkmalspunkte zeilenweise verschiedenfarbig gekennzeichnet und die Entzerrung mit den ermittelten Verzerrungsparametern durchgeführt worden. Die ermittelten Verzerrungsparameter lauten bei der verwendeten Kamera Nr. 1 aus Tab. 4.2 (Auflösung $1024\text{px} \times 768\text{px}$) mit dem Tamron-Objektiv ($f = 4\text{mm}$) und einem hinzugefügten Rahmen von 78px :

- $f'_x = f'_y \approx 882,0 \text{ px}$,
- $c_x \approx 589,5 \text{ px}$, $c_y \approx 461,5 \text{ px}$,
- $k_1 \approx -0,38$, $k_2 \approx 0,16$,
- $p_1 \approx -0,00024$, $p_2 \approx 0,00031$.

Da $f'_x = f'_y$ ist, sind die Pixel quadratisch. Die gegenläufigen radialen Verzerrungsparameter sind in Abb. 3.7 gut an der Form der Ränder zu erkennen: k_1 wirkt zur Bildmitte stärker, k_2 wirkt dem zu den Ecken hin entgegen. Der Einfluss der kleinen tangentialen Parameter ist kaum wahrnehmbar.

Mit den gleichen Verzerrungsparametern ist die Entzerrung von Abb. 3.2 a zu b durchgeführt worden. Die äußere Orientierung wird hier durch das manuell positionierte blaue Koordinatensystem in Abb. 3.2 c festgelegt. Dieses liegt mit der Aufhängenhöhe von $a \approx 537 \text{ cm}$ bei

$$\vec{t} \approx \begin{pmatrix} 2 \\ 101 \\ 537 \end{pmatrix} \text{ cm}.$$

Die Rotationsmatrix ist ähnlich der Einheitsmatrix, da die Blickrichtung orthogonal festgelegt ist und die sonstige Ausrichtung zum Versuchsaufbau nur leicht um $\alpha \approx -3^\circ$ um die z-Achse gedreht werden muss:

$$R = \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Auf dem Boden ist der Skalierungsfaktor gemäß Gleichung 3.6

$$s = a/f' \approx 0,6 \text{ cm/px}.$$

In Kopfhöhe liegt das Pixel-Meter-Verhältnis aufgrund der perspektivischen Sicht ungefähr bei $0,4 \text{ cm/px}$.

3.2 Detektion mit Marker

Wie in der Einleitung beschrieben, haben wir Versuche durchgeführt, um verlässliche Daten der Fußgängerdynamik zu erhalten. Erst mit Laborversuchen ist man in der Lage, wohldefinierte Rahmenbedingungen zu schaffen und nur relevante Parameter zu variieren. Für die robuste und exakte Detektion war es daher möglich, die Probanden mit Markern auszustatten. Da wir mit optischen Systemen und zum Teil hohen Dichten gearbeitet haben, wurden die Marker auf dem Kopf befestigt und die Aufnahmen überkopf durchgeführt. Die Nutzung des ggf. positionsstabileren Mittelpunktes zwischen markierten Schultern [325] hätte hier insbesondere zum Bildrand zu vielen Verdeckungen und damit unerkannten Personen geführt. Dabei kamen je nach Randbedingungen unterschiedliche Marker zum Einsatz (siehe Abb. 3.8 und 3.9). Diese Randbedingungen können wie folgt zusammengefasst werden:

1. Pixel-Meter-Verhältnis: abhängig von möglicher Aufhängenhöhe, Abdeckungsbereich, Sensorauflösung (welche wiederum abhängig von der Bildwiederholrate und der Übertragungsrate ist) (siehe Abschnitt 4.1.1.3, 4.1.1.5, 4.1.1.7 und 4.1.1.9),

2. Verdeckung, Überlappung: abhängig von maximaler Dichte, Größenunterschieden, Geometrie (Aufbauten, Bodenbeschaffenheit wie Treppen) (siehe Abschnitt 4.1.1.1 und 4.1.1.2),
3. benötigte Zusatzinformationen: z. B. Blickrichtung, Individualisierung.

Im Folgenden sollen die eingesetzten Marker aus Abb. 3.8 und 3.9 vorgestellt und die Eigenschaften erläutert werden, wobei diese immer auf dem Hintergrund der mit diesen Markern durchgeführten Experimente zu sehen sind. Die Gründe für die Wahl bzw. Anpassung sind ebenfalls der Aufzählung zu entnehmen und die Eigenschaften gesammelt in Tab. 3.1 zu finden. Die Tabelle zeigt die unterschiedlichen Stärken und Schwächen der einzelnen entwickelten Marker. Die Experimente und deren Umsetzung werden ausführlich in Kapitel 4 beschrieben. Der Einsatz der Marker in den Versuchen kann Tab. 4.1 entnommen werden. Die Namen der hier durchnummerierten Marker sind in der Software *PeTrack* zum Teil nach der Versuchsreihe, für die sie anfänglich entwickelt wurden, benannt: Marker Nr. 1 ist der Kasernen-, Nr. 2 der Hermes-, Nr. 3 der Japan-, Nr. 4 der Farb- und Nr. 5 der Codemarkers.

Marker 1 ist komplex und damit robust in der Detektion. Die Anzahl von falsch positiv detektierten Personen (erkannte Marker sind keine Marker bzw. erkannte Personen keine Personen) sind daher nahezu Null (für alle Versuche der Versuchsreihe Nr. 3 kleiner als 0,2%). Die Anzahl von falsch negativen Detektionen (Marker wurden nicht erkannt) sind aufgrund hoch angesetzter Akzeptanzschwellen je Bild hoch (ca. 5%), haben wegen des verbindenden Trackings jedoch einen geringeren Einfluss auf das Gesamtergebn, da dies am Ende nur zu einem Fehler führt, wenn die Person während des gesamten Durchquerens des erfassten Versuchsbereichs nicht erkannt wird (siehe später in diesem Abschnitt). Im vollständigen Video nicht erkannte Personen kamen selten vor (für alle Versuche der Versuchsreihe Nr. 3 kleiner als 1%, vgl. Abschnitt 3.6.3). Die Funktionsweise der Erkennung für diesen Marker wird später im Abschnitt erläutert. Marker 1 ermöglicht das Auslesen der Körpergröße und der Kopfausrichtung, nicht jedoch der Blickrichtung, der erkannten Person. Welche Unterschiede es im Umfang der mit den Trajektorien verknüpften Daten gibt, kann Abschnitt 2.1 entnommen werden (Markernutzung in Video Nr. 50).

Marker 2 hat aufgrund der einfachen Struktur die größte Fehlerrate (siehe Abschnitt 3.6.3), sowohl bei der Anzahl von falsch positiven als auch falsch negativen Detektionen, insbesondere bei den Versuchen an Treppen im Freien (siehe Abb. 4.8). Der Marker wurde trotzdem so gewählt, da bei den Überkopfaufnahmen an Treppen die Personen mit ihren Köpfen nah beieinander lagen (siehe unterer Bereich des Bildes der Spalte c in Abb. 3.8), so dass der Einsatz größerer glatter Flächen wie bei Marker 3 nicht möglich war. Die große Entfernung in den Tribünen von über 13 m und die nicht fokussierbaren Optiken der eingesetzten Stereokameras, die keine optimal scharfen Abbildungen erlaubten, bedurfte jedoch eines größeren Punktes als bei Marker 1, so dass minimal $3\text{ px} \times 3\text{ px}$ gemäß Gleichung 4.7 (Einheit Pixel sensorseitig und Kameraentfernung in durchschnittlicher Kopfhöhe von 176 cm) und Tab. 4.2 zur Erkennung existierten:

$$p_m = p_h \frac{w_m}{l_h} = p_h \frac{w_m}{2h_k \tan \alpha} = 1280\text{ px} \frac{4\text{ cm}}{2(1475\text{ cm} - 176\text{ cm}) \tan 32^\circ} \approx 3\text{ px}.$$

Den gleichen Erfassungsbereich mit mehr Kameras mit Objektiven größerer Brennweite abzudecken, war aufgrund der Nutzung teurer Stereokameras wegen der Versuche an Treppen nicht finanzierbar. Da der Marker direkt auf die eng anliegende Kopfbedeckung (Bandana) gedruckt

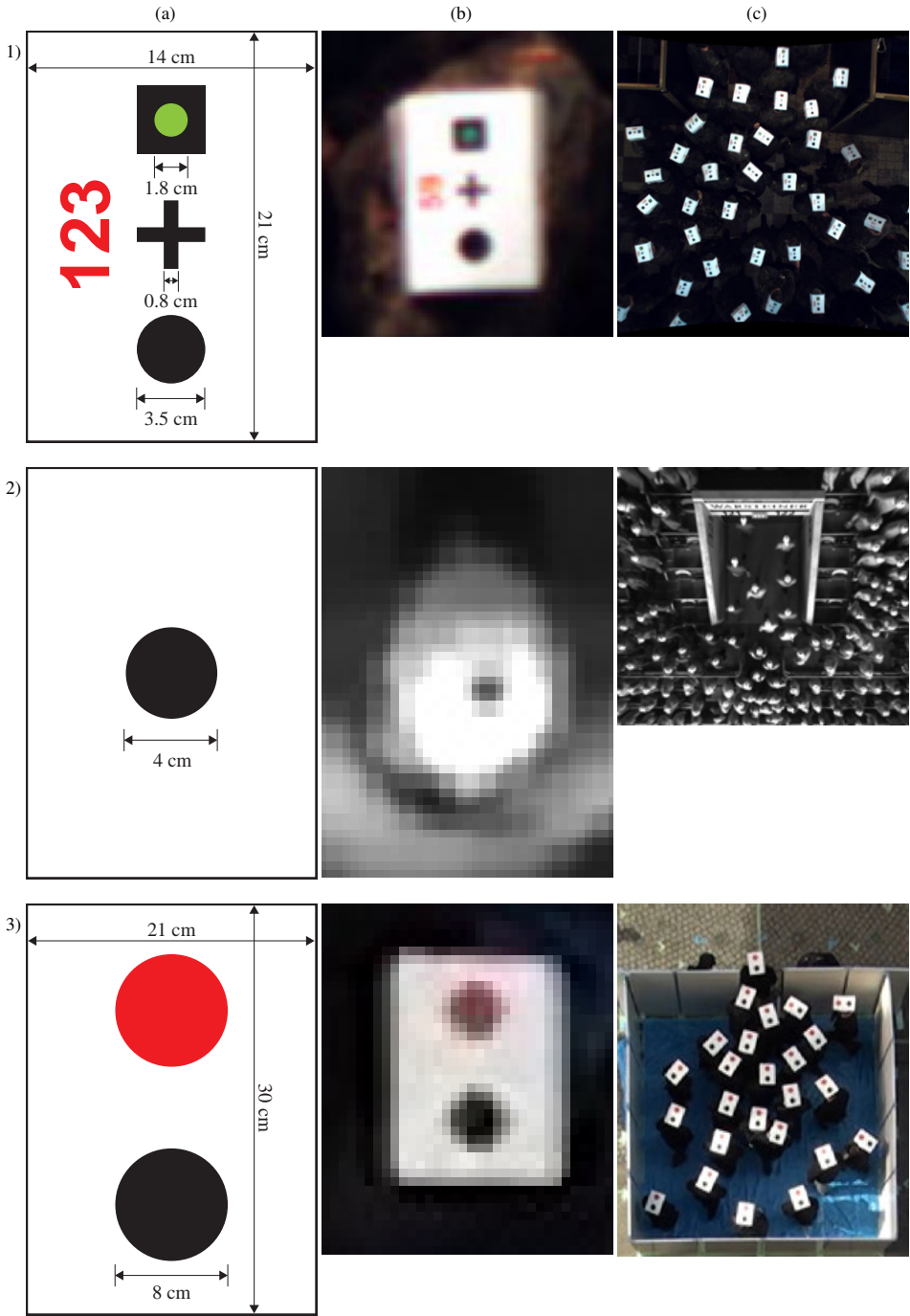


Abb. 3.8: Liste bisher eingesetzter Marker; weiteres siehe folgende Abb. 3.9

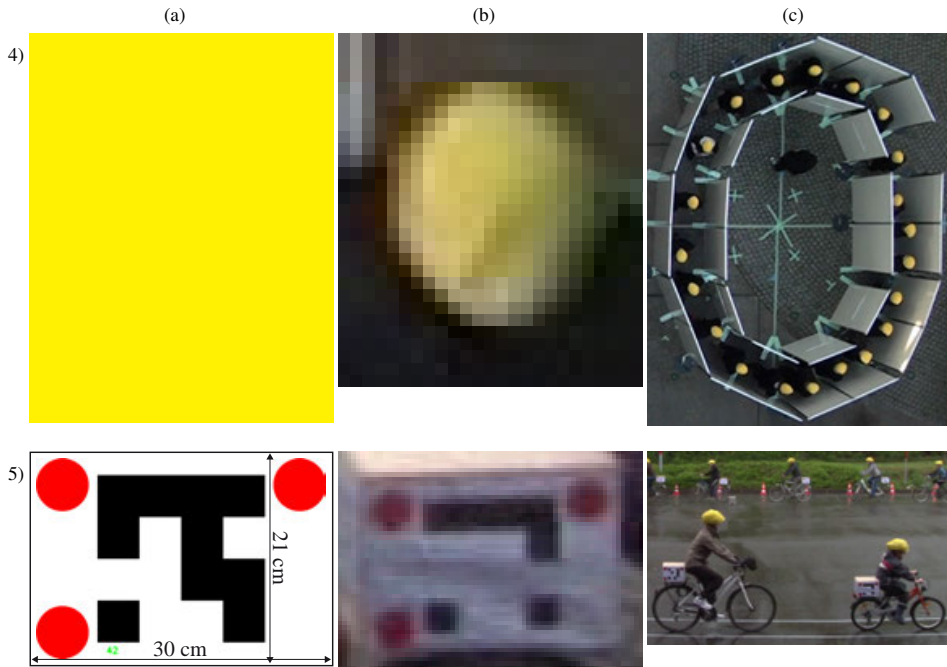


Abb. 3.9: Liste bisher eingesetzter Marker (siehe auch Abb. 3.8); bei welchen Versuchsreihen welche Marker eingesetzt wurden, kann Tab. 4.1 entnommen werden; Spalte (a) zeigt den Marker, Spalte (b) eine Aufnahme des Markers bei einem Versuch und Spalte (c) einen Versuch in Gänze

wurde, war auf dem Kopf kein Platz für weitere Elemente, die auch bei leicht unterschiedlicher Kopfhaltung und Sitz der Kopfbedeckung zu sehen gewesen wären (siehe Spalte b in Abb. 3.8). In den Tribünen erschwerte zudem die stark schwankende Sonneneinstrahlung bei offenem Dach die Erkennung, die mit Linien gleicher Helligkeit arbeitet und daher einer gleichmäßigen Ausleuchtung bedarf. Die Erkennungstechnik ist der Erläuterung zu Marker 1 später in diesem Abschnitt zu entnehmen. Der Farbmarker 4 konnte für die Hermes-Versuche (Versuchsreihe Nr. 6 in Tab. 4.1) nicht genommen werden, da zum einen die Köpfe an den steilen Treppen in den Tribünen sich teilweise überlappten und damit als ein Farbbobjekt wahrgenommen worden wären und zum anderen die Sitze in den Tribünen der ESPRIT Arena in Düsseldorf unterschiedlich farbig sind (Markernutzung in Video Nr. 44).

Marker 3 wurde eingeführt, da die Eigenschaften des Markers 1 wie die Robustheit und die Bestimmbarkeit der Kopfausrichtung gewünscht, aber die Pixelabdeckung bei den durchgeführten Versuchen kleiner war (siehe Spalte b in Abb. 3.8). Daher mussten die Pappe und auch die Markerelemente doppelt so groß gewählt werden, was in dichten Situationen zu falsch negativen Detektionen in einigen Bildern führte. Durch das verbindende Tracking konnten aber ähnlich geringe Fehlerraten wie bei Marker 1 erzielt werden, obwohl auch hier im Freien gearbeitet wurde und starke Helligkeitsunterschiede existierten (siehe Spalte c in Abb. 3.8). Jedoch war durch die Nutzung einer ebenen Pappe wie bei Marker 1 (siehe Abb. 4.15) die Helligkeitsverteilung auf der einzelnen Pappe homogen und somit Linien gleicher Helligkeit identifizierbar (Markernutzung in Video Nr. 28).

Marker 4 bedarf der wenigsten Vorbereitung. Eine eng anliegende Mütze in einer Farbe, die im Detektionsbereich nicht vorkommen darf, dient der Erkennung. Die wichtigste Voraussetzung hierbei ist neben der Alleinstellung der Farbe die Ausleuchtung der Versuchsfläche mit nur einer Farbe bzw. Farbtemperatur. Die größten Nachteile dieses Markertyps sind bei der jetzigen Implementation, dass mit ihren Köpfen sich zu nah kommende Personen zu einem Marker bzw. einer Person verschmelzen (für die nötige Einschränkung des Bildwinkels siehe Abschnitt 4.1.1.1) und dass bei der Annahme des Detektionsortes mit der Mitte des Farbbereichs ein Fehler u. a. in Abhängigkeit vom Blickwinkel einhergeht (siehe Abschnitt 3.6.4.2). Zu differenzierende Gruppen könnten mit unterschiedlich farbigen Mützen ausgestattet werden. Eine genaue Erläuterung des Verfahrens folgt im Unterabschnitt 3.2.2 (Markernutzung in Video Nr. 27).

Marker 5 hat die größte Komplexität und damit zu vernachlässigende Fehler falsch positiv detektierter Personen. Die Komplexität dient der Individualisierung und soll so jede Person identifizierbar machen. An dieser Stelle können je nach Anzahl von zur Fehlerkorrektur eingebrachten Kontroll-, Prüf- bzw. Paritätsbits (*parity bits*) Fehler beim Auslesen gemacht werden, die aber wie bei den vorangegangenen Markern durch das verbindende Tracking verringert werden können. Abb. 3.9 zeigt den Einsatz bei Fahrradexperimenten parallel zu den Farbmarkern 4. Aussagen in dieser Arbeit (siehe u. a. Abschnitt 4.1.1) beziehen sich jedoch auf den alleinigen Einsatz als Kopfmarker bei den BaSiGo-Versuchen (Nr. 9), bei deren Erfassung Kameras mit einer großen Brennweite genutzt werden müssen, um die nur 5 cm × 5 cm großen Codes erfassen zu können. Die Erfassung bzw. das Auslesen des Codes selber ist nicht Gegenstand dieser Arbeit. Die Individualisierung ermöglicht mit der Befragung der Probanden eine noch tiefergehendere Analyse der Laufwege. Abhängigkeiten der Bewegung z. B. vom Geschlecht, der Fitness, der Körperproportionen oder der Gruppenzugehörigkeit (vgl. [462]) sind so erstmals differenzierbar (Markernutzung in Video Nr. 21 und 9).

Marker Nr.	Robustheit der Detektion	Genauigkeit der Position	hohe Dichte	zusätzliche Info.	benötigte Auflösung	Aufwand der Herstellung
1	++	++	o	+	o	o
2	o	++	+	-	+	+
3	+	++	-	o	+	o
4	+	o	-	o	++	++
5	++	++	+	++	-	-

Tab. 3.1: Eigenschaften der verwendeten Marker in der verwandten Ausprägung (Codemarker (Nr. 5) bei Anwendung in BaSiGo-Versuchen (Nr. 9)) (++) sehr gut, + gut, o mäßig, - schlecht

Die zur Annahme eines Markers erforderlichen Eigenschaften wurden generell hoch angesetzt, da eine Person nicht in jedem Bild erkannt werden muss, sondern das Tracking (siehe Abschnitt 3.3) die Trajektorien weiterführt. Um für eine erst spät erkannte Person über den ganzen Sichtbereich bzw. die Zeit vor der Erkennung eine durchgängige Trajektorie zu haben, wird das Tracking auch zeitlich rückwärts durchgeführt. Bei der manuellen Kontrolle ist zudem in dichten Menschenmengen eine gar nicht erkannte Person besser zu erkennen, als eine Fehldetektion.

In dieser Arbeit ist mit Erkennung oder Identifizierung einer Person im Weiteren nur die Lokalisierung einer unbestimmten Person gemeint, bei der nur sehr beschränkt Eigenschaften der entsprechenden

Person wie die zugehörige Größenklasse oder Blickrichtung ausgelesen werden können. Nur der Codemarker (Nr. 5) erlaubt eine Unterscheidung der Trajektorien für eine Individualisierung der Laufwege. Durch Abgleich des ausgelesenen Codes mit den ausgefüllten Fragebögen der Teilnehmer können darüber alle relevanten Informationen mit den Laufwegen verknüpft werden.

Der für die Erkennung relevante Bildbereich (*region of interest*, ROI) kann eingegrenzt werden und wird z. B. in Abb. 3.2 d durch ein grünes Rechteck angezeigt. Dieses ist in PeTrack manuell einstellbar und sollte so gewählt werden, dass diesen Bereich alle zu detektierenden Personen durchlaufen und die Ausleuchtung die Marker mit möglichst wenig Fehlern erkennen lässt. Der Bereich sollte möglichst groß sein, um die Personen möglichst oft erfassen zu können, so dass für eine Szene ein Optimum zwischen Bereichsgröße und robuster Erkennung gefunden werden muss. Die Laufwege zeitlich wie örtlich vor und hinter dem eingegrenzten Bereich werden mittels des Trackings erfasst (siehe Abschnitt 3.3). Die Eingrenzung beschleunigt zudem den Erkennungsprozess.

Vor dem Erkennungsschritt können in PeTrack einige Filter auf das gesamte Bild angewandt werden, um die Detektionsgüte zu erhöhen. Hierzu gehören u. a. die Anpassung der Helligkeit oder des Kontrastes. Zudem ist eine Maskierung bzw. Subtraktion des Hintergrundes (*background subtraction*) implementiert (siehe Abb. 3.10). Dabei wird ein robustes, schnell lernendes Verfahren [463] verwandt, das auch Schatten als Hintergrund erkennt und für die nachfolgende Erkennung die im Video nicht verändernden Bereiche maskiert. Es wird dabei angenommen, dass sich dort keine Personen befinden können. Diese Annahme und die Anwendung der Hintergrundsubtraktion ist nur erfolgreich, wenn der Hintergrund lokal die meiste Zeit zu sehen ist und insbesondere eine nicht dichte, sich bewegende Menschenmenge zu untersuchen ist. Bei der Arbeit mit markierten Personen kann dieser rechenzeitintensive Schritt jedoch deaktiviert werden, da die Anzahl der Fehldetektionen gering ist und bei den oftmals hohen Dichten die Hintergrundbestimmung fehleranfällig ist. In Abschnitt 3.5.1 wird ein selbst implementierter Ansatz beschrieben, der bei der Detektion nicht markierter Personen mittels Stereokameras zum Einsatz kommt. Dort ist die Hintergrundsubtraktion zur Vermeidung von Fehldetektionen essentiell.

Die Erkennung der Marker soll im Folgenden exemplarisch am Kasernen-Marker (Nr. 1) und Farb-Marker (Nr. 4) erläutert werden.



Abb. 3.10: Anwendung der Hintergrundsubtraktion; der bestimmte Hintergrund ist ausgegraut und zeigt, dass nur der sich sehr vom Hintergrund abhebende Marker als Vordergrund angenommen wird

3.2.1 Marker mit Strukturelementen

Marker 1 wurde auf eine DIN-A4-Pappe gedruckt und an den kurzen Seiten so geknickt und am Hut der Soldaten bei den Kasernen-Versuchen (Nr. 3) mit Büroklammern befestigt, dass die nach oben sichtbare Fläche 14 cm × 21 cm und damit ungefähr kopfgroß war. Auf den Seiten und oben war eine individuelle Nummer aufgebracht, um die Probanden sortieren oder ansprechen zu können. Die für die Erkennung relevanten drei Struktur- bzw. Markerelemente befinden sich auf einer Geraden ent-

lang der Kopfausrichtung: ein schwarzer Kreis, ein einen Farbkreis beinhaltendes schwarzes Quadrat und ein schwarzes Kreuz. Das Kreuz wird als Position der Person angenommen und beschreibt damit den Laufweg, wobei immer mit subpixelgenauen Werten gearbeitet wird. Das Quadrat zeigt die Gesichtsseite an.

Marker 1 wurde bei den Kasernen-Versuchen (Nr. 3) eingesetzt. Dort kam u. a. Kamera Nr. 1 mit unterschiedlichen Objektiven zum Einsatz. Vor der Entzerrung lag je nach Objektiv und Versuch die minimale Pixelabdeckung in den äußeren Ecken in durchschnittlicher Kopfhöhe bei 1,59 px/cm und maximal in der Bildmitte bei 2,59 px/cm. Die großen Unterschiede rühren von der starken Objektivverzerrung her (siehe Abb. 3.7). Abb. 3.8 und 3.9 zeigen die Marker optimal scharf, da sich die Personen zum Zeitpunkt der Aufnahmen kaum bewegten. Bei der Bildskalierung wurde in Spalte b keine Pixelinterpolation durchgeführt, so dass die Spalte b einen guten Eindruck für die Anzahl an Pixeln der einzelnen Markerelemente gibt: in der Bildmitte ca. 35 px \times 53 px für die gesamte sichtbare Pappe und ca. 4 px \times 4 px für den Farbpunkt. Für Marker 1 zeigt sich die verhältnismäßig gute Auflösung, da die Kameras nur ca. 350 cm über den Köpfen angebracht waren.

Die Blende (wenn möglich) und Belichtungszeit wurden so eingestellt, dass die Marker möglichst kontrastreich erschienen (siehe Abb. 3.8, Spalte c).

Die eigentliche Erkennung basiert für alle Marker mit Strukturelementen in Abb. 3.8 auf der Größe und Lage von Linien gleicher Helligkeit. Abb. 3.11 und 3.12 für einen Ausschnitt zeigen diese Isolinien in der linken Spalte. Dabei kommen gerichtete Isolinien zum Einsatz, die entweder einen helleren Bereich (rote Linien) oder dunkleren Bereich (grüne Linien) einschließen. Die Abbildungen zeigen vier der angewandten Helligkeitswerte. Es werden insgesamt zehn Helligkeitswerte verwandt, die mittels der Software *PeTrack* eingestellt werden können. Dass ein großer Wertebereich für die Helligkeit nötig sein kann, ist ebenfalls den Abbildungen zu entnehmen, in denen die Pappe je nach Kopfneigung und Ausleuchtung unterschiedlich hell erscheint. Wenn die Markerelemente wie bei Versuchsreihe Nr. 3 und 8 auf einer ebenen Fläche aufgebracht sind, sind die Helligkeitswerte gleicher Farben ähnlich. Dies ist für die Arbeit mit Isolinien gleicher Helligkeit von Vorteil.

Alle Isolinien werden durch Ellipsen approximiert (siehe rechte Spalten der Abb. 3.11 und 3.12), um einen einfachen Zugriff auf die Form (u. a. Seitenverhältnis), Orientierung, Größe und Schwerpunkt bzw. Position der Isolinien zu haben. Die Approximation der Isolinien durch Ellipsen geschieht unter Anwendung der kleinsten Fehlerquadratsumme zu den Punkten, die die Geradensegmente der Isolinie beschreiben [464]. Dieser Ansatz kann gewählt werden, da die Isolinien für gewöhnlich keine geraden Kanten aufweisen (siehe Abb. 3.13 a) und damit die die Isolinien beschreibenden Punkte der Geradensegmente dicht liegen.

Im Anschluss werden alle Ellipsen zusammengefasst, die das gleiche Objekt umschreiben. Separiert nach ihrer Richtung wird dabei die Ellipse, die eine kleinere zugehörige Fläche umschreibt, und deren Mittelpunkt innerhalb einer Ellipse liegt, die eine größere Fläche umschreibt, dieser zugeordnet. Von den zusammengefassten Ellipsen wird diejenige am Ende ausgewählt, die von ihrer Größe und dem Seitenverhältnis dem theoretischen Optimum der Form des realen Markerelementes am besten entspricht, je nachdem ob es sich um eine die ganze Pappe umschreibende Ellipse (rot) oder um eine die einzelnen Markerelemente umschreibende Ellipse (grün) handelt (siehe Abb. 3.13 b). Das Seitenverhältnis der Halbachsen der roten und damit den ganzen Kopf umschreibende Ellipse sollte annähernd 3:2 betragen, was einer numerischen Exzentrizität von $\sqrt{5}/3$ entspricht, und der grünen Ellipsen 1:1, was einer numerischen Exzentrizität von 0 entspricht. Die grünen Markerellipsen werden den roten Kopfellipsen zugewiesen, in dessen Fläche der Mittelpunkt der Markerellipse liegt.

Am Ende werden maximal die vier Markerellipsen eines Kopfes genutzt, die die dunkelsten Bereiche

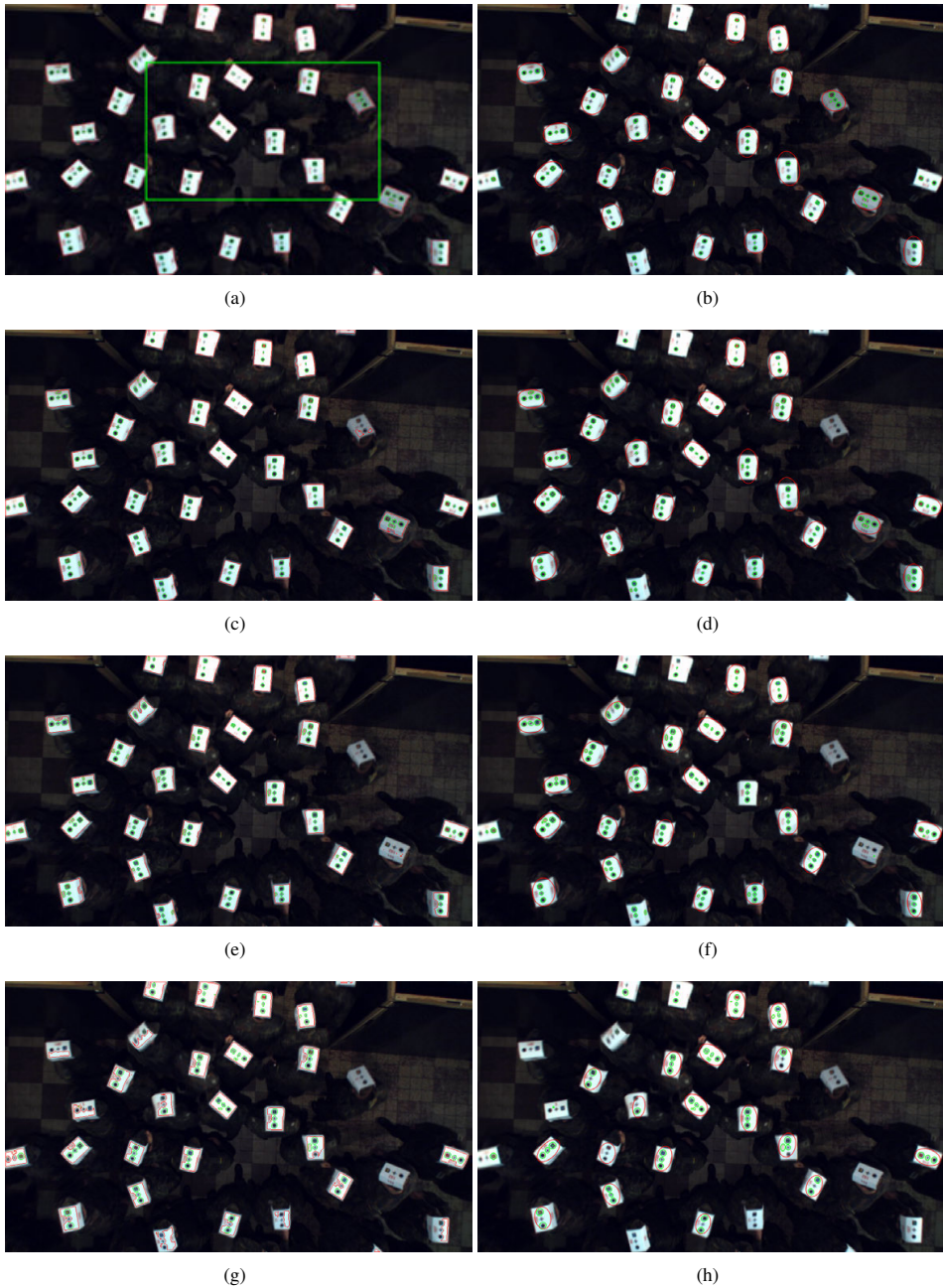


Abb. 3.11: (links) gerichtete Isolinien gleicher Helligkeit (rot: von außen dunkel nach innen hell; grün: von außen hell nach innen dunkel) mit Helligkeitswerten von (a) 110, (b) 144, (c) 178 und (d) 229 $\in [0, 255]$; des relevanten Bildbereichs im grünen Rechteck aus Abb. 3.2 d; (rechts) interpolierende Ellipsen der Isolinien des entsprechenden linken Bildes; der grün umrandete Bereich in (a) wird für alle Bilder in Abb. 3.12 vergrößert dargestellt

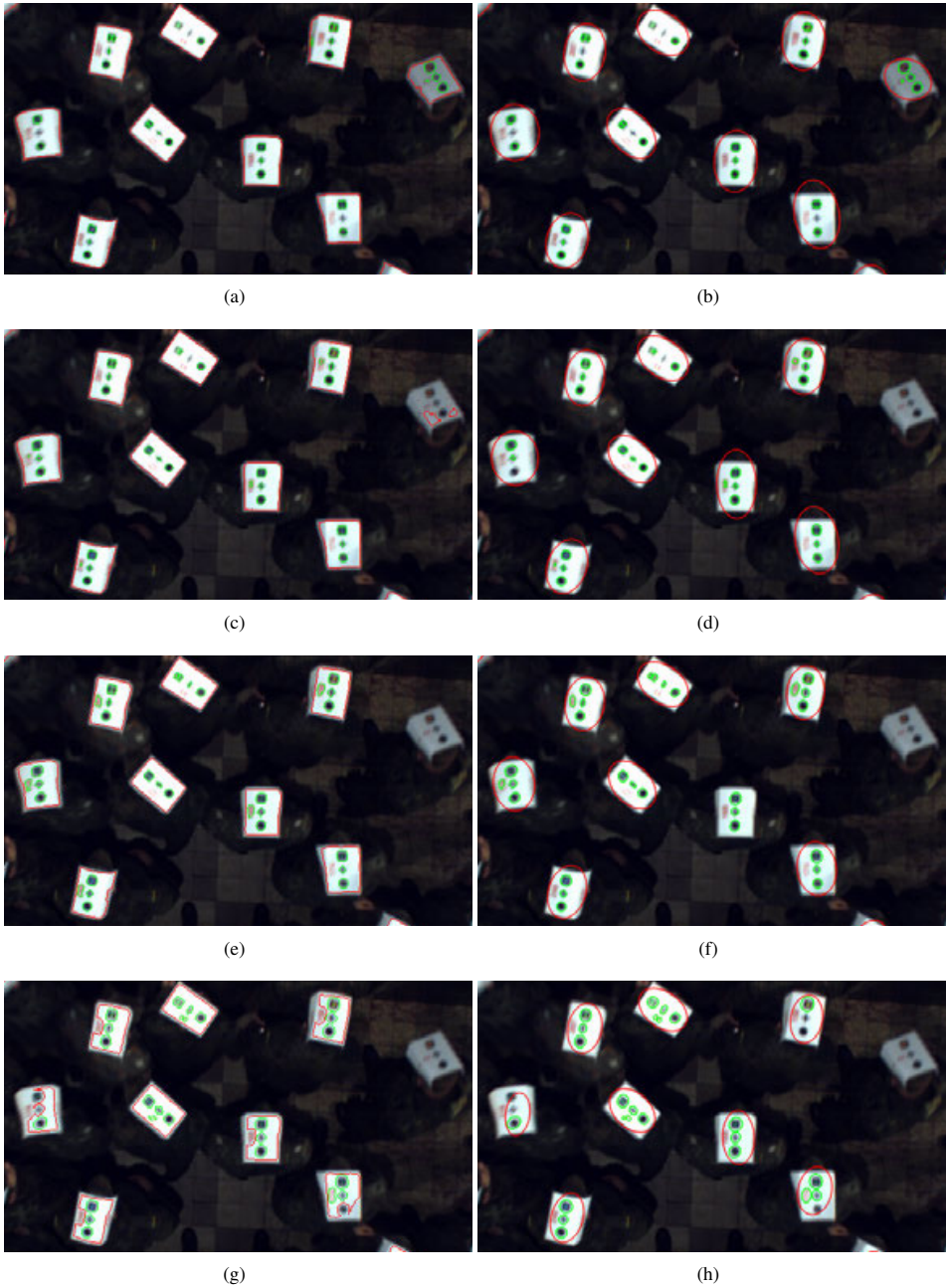


Abb. 3.12: Ausschnitt der entsprechenden Bilder aus Abb. 3.11

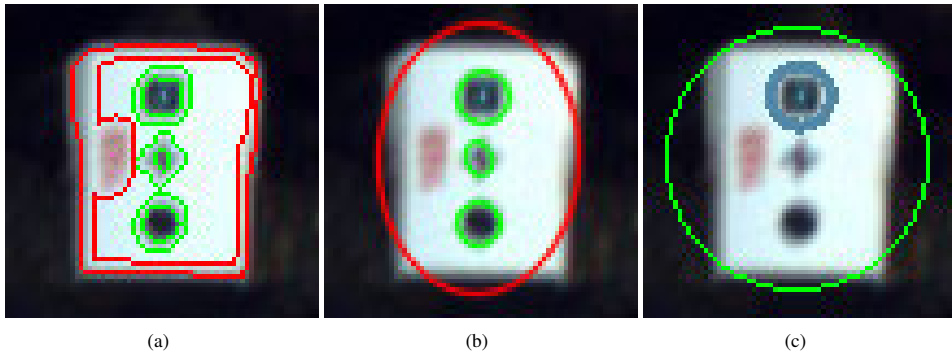


Abb. 3.13: Schritte zur Erkennung des Kasernen-Markers (Nr. 1); (a) gerichtete Isolinkien gleicher Helligkeit für zwei Helligkeitswerte; (b) von allen approximierenden Ellipsen die je nach Art optimalen; (c) Visualisierung in PeTrack; Farbmarkerelement und Kopfposition werden mit einem passend eingefärbten und der Kopfgröße angepassten Kreis (abhängig von Körpergröße bzw. in 3D vom Abstand zur Kamera) angezeigt

einfassen. Vier Elemente sind nötig, da auch die rote Zahl auf dem Kopf aufgebracht ist und manchmal dunkler als ein Markerelement erschien. Von diesen maximal vier Markerellipsen werden drei bestimmt, deren Mittelpunkte fast auf einer Geraden liegen und einen ähnlichen Abstand zueinander besitzen. Die Gerade zeigt die Kopfausrichtung und die größte Markerellipse mit den höchsten Helligkeitswerten nahe ihres Mittelpunktes die Vorderseite des Kopfes an. Dabei umschließt die größte Markerellipse das Farbmarkerelement. Der Mittelpunkt der mittleren Markerellipse kennzeichnet die Position des Kopfes und damit der zu verfolgenden Person.

Ist ein Aufnahmebereich so überbelichtet, dass die einzelnen Markerelemente nicht oder kaum noch zu erkennen sind, besteht in PeTrack die Möglichkeit, eine Person nur über die Kopfellipse zu detektieren. Als Kopfposition wird dann der Mittelpunkt der Kopfellipse verwendet. Die Position dieses Mittelpunktes liegt entlang eines Laufweges nicht so stabil an der gleichen Stelle wie der Mittelpunkt der Ellipse des zentralen Markerelementes.

Es besteht zudem in PeTrack die Möglichkeit, für Marker mit einem dunklen Zentrum (Marker Nr. 1 und 2) den zu verfolgenden Punkt auf Subpixel-Ebene zu verbessern, indem die dunkelste Stelle im Zentrum bestimmt wird.

Die Farbe des Punktes auf dem quadratischen Markerelement wird durch das hellste Pixel in der Nähe des entsprechenden Ellipsenmittelpunktes bestimmt (für die Kasernen-Versuche (Nr. 3) wird ein Bereich von $3\text{px} \times 3\text{px}$ um den Ellipsenmittelpunkt abgesucht). Abb. 3.13 c zeigt die ermittelte Farbe als Kreisfarbe um das entsprechende Markerelement. Die am Ende einer Person zugeordnete Farbe ist der Mittelwert aller detektierten Farben entlang des Laufweges, wie in Abb. 3.2 f zu sehen ist. Die Nutzung der Farbe als Größenindikator wird in Abschnitt 3.4.1 beschrieben.

Die Erkennung geschieht wie beschrieben interframe und nutzt keine darüber hinausgehenden intra-frame Informationen. Für die Implementierung eines neuen Markers ist damit in PeTrack lediglich der Detektionsschritt zu ersetzen. Das nachfolgende Tracking geschieht unabhängig vom Erkennungsschritt.

3.2.2 Farbmarker

Der Farbmarker ist am einfachsten in der Herstellung bzw. Beschaffung, da käuflich zu erwerbende farbige Mützen genutzt werden können. Er kann aber trotzdem bei geeigneter Farbwahl gut detektiert werden. Eine Person wird dann angenommen, wenn eine zusammenhängende Fläche eine ähnliche Farbe besitzt und dabei die abgedeckte Pixelfläche und das Seitenverhältnis der Ausdehnung in einem definierten Rahmen liegt. Die genaue Position ist dabei der Mittelpunkt der die Umrandung der Farbfläche approximierenden Ellipse. Das sonstige Äußere der Versuchsteilnehmer und die Umgebung sollten die zur Markierung gewählte Farbe nicht enthalten, zumindest nicht in Größe und Form der Markerfläche.

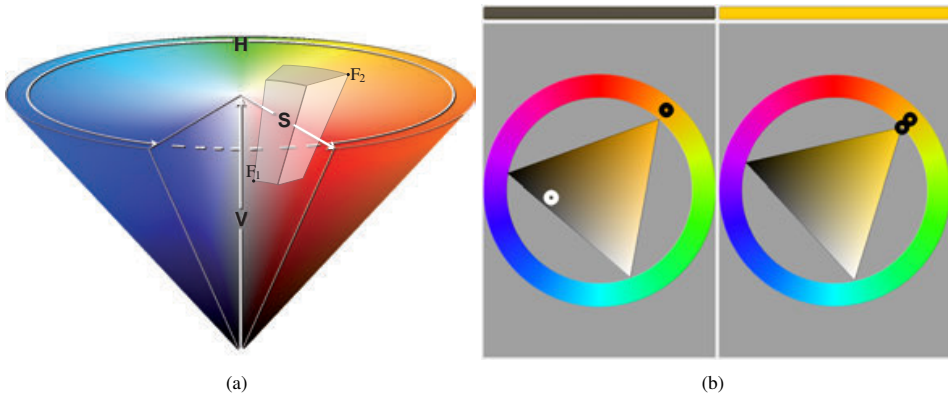


Abb. 3.14: (a) Kegel des HSV-Farbraumes mit dem Farbton H, der Sättigung S und der Helligkeit V sowie einem eingeschriebenen grauen Unterraum, der über zwei Farben F_1 und F_2 angegeben werden kann; (b) Möglichkeit der Farbwahl von F_1 und F_2 in PeTrack; Bild a basiert auf Abbildung in [465]

Die bei einer matten, einfarbigen Kopfbedeckung innerhalb einer Aufnahme auftretenden Farbvariationen rühren insbesondere von unterschiedlichen Einfallswinkeln des Umgebungslichtes her. Gemäß Lambertschem Gesetz [466] fällt bei der Annahme einer ideal diffusen Reflexion die Intensität bzw. Helligkeit der ausgestrahlten Farbe mit steigendem Winkel α zwischen der Oberflächennormalen und dem Winkel des einfallenden Lichtes um den Faktor von $\cos \alpha$ ab. Die Kopfbedeckung muss dabei matt bzw. nicht reflektierend sein, da ansonsten auch die Eigenfarbe des Lichtes auf der Kopfbedeckung zu sehen sein würde [467]. Um aufgrund der reinen Helligkeitsunterschiede der Farbmarker die Farbe getrennt nach Farbton und Helligkeit eingrenzen zu können, kann in PeTrack der Wertebereich der für die menschliche Wahrnehmung dreidimensionalen Farbe im HSV-Farbmodell [468] angegeben werden. Abb. 3.14 a zeigt den durch die drei Primärvalenzen H (*hue* = Farbton), S (*saturation* = Sättigung) und V (*value* = Helligkeit) aufgespannten kegelförmigen Farbraum und ein graues Teilvolumen, das als Farbbereich durch die Farben F_1 und F_2 der Eckpunkte beschrieben werden kann. In PeTrack können diese beiden Farben, u. a. wie in Abb. 3.14 b zu sehen, festgelegt werden. Die hier gewählten Farben $[H, S, V]$ sind $F_1 = [37, 28, 72]$ und $F_2 = [48, 255, 255]$, wobei im äußeren Kreis der Farbton und im inneren Dreieck die Sättigung und Helligkeit festgelegt wird. Die in der Abb. 3.14 b gewählten Farben wurden für die Detektion in Abb. 3.15 herangezogen.

Die komplette grafische Benutzerschnittstelle zur Einstellung der gesamten Optionen des Farbmarkers kann Abb. D.4 entnommen werden. Neben der Angabe eines Wertebereichs für die Pixelanzahl und des maximalen Seitenverhältnisses der approximierenden Ellipse kann die Größe des Strukturele-

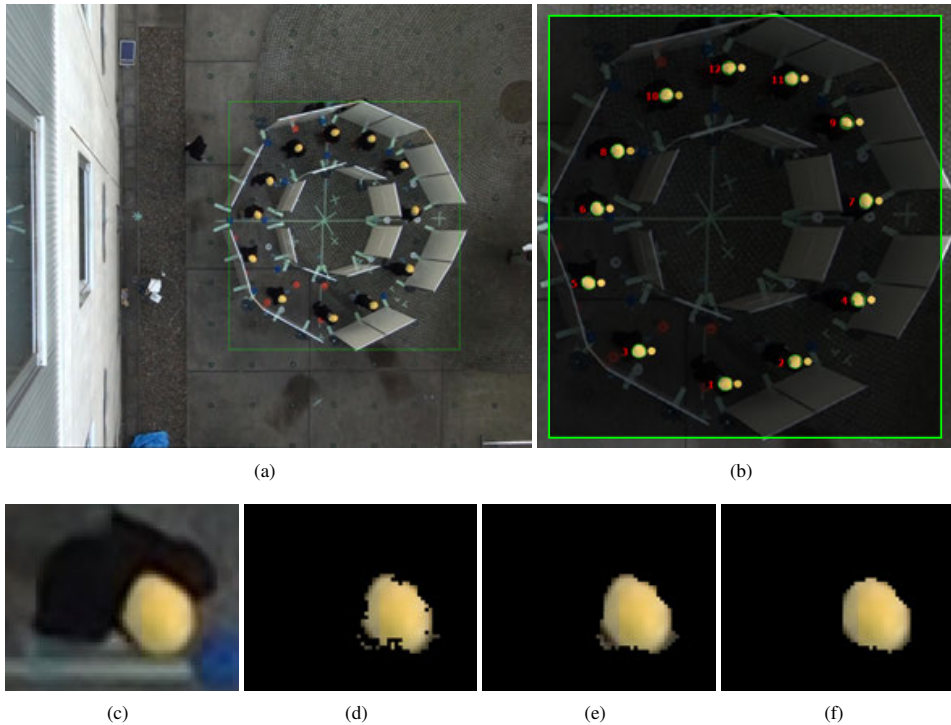


Abb. 3.15: (a) Experiment der Japan-Versuche (Nr. 8) mit Nutzung des Farbmarkers (gelbe Mützen) (Nr. 4); (b) zeigt alle über die farbige Kopfbedeckung detektierten Personen innerhalb des grünen Rechtecks (ROI); die nicht zur selektierten Farbe passenden Pixel sind ausgegraut und die erkannte Farbe einer jeden Person und eine fortlaufende Nummer wird dargestellt; (c) ist ein Bildausschnitt einer Person, dessen Pixel in (d) maskiert sind, die nicht im Bereich $[F_1, F_2]$ aus Abb. 3.14 b liegen; in (e) wurde ein Schließen und in (f) ein Öffnen der Maske mit einem Kreis vom Durchmesser 5 px durchgeführt

menten (*kernel*) für das Öffnen und Schließen der Maske zur Farbwahl angegeben werden. Die morphologische Basis-Operation des Schließens (*closing*) [469] ermöglicht u. a. das Herausfiltern lokal begrenzter Störungen. Abb. 3.15 c zeigt eine zu detektierende Person des in Abb. 3.15 a abgebildeten Experiments. In d ist der Bereich maskiert (schwarz überdeckt), der nicht im eingestellten Farbintervall $[F_1, F_2]$ liegt. Abb. 3.15 e zeigt nun die mittels eines Kreises mit dem Durchmesser 5 px als strukturierendes Element geschlossene Maske, bei der Störungen im Inneren wie auch am Rand verringert wurden. Das Schließen entspricht dabei der nacheinander ausgeführten Dilatation und Erosion jeweils mit demselben strukturierenden Element. Durch die Dilatation werden alle Löcher geschlossen, in die das strukturierende Element nicht vollständig hineinpasst. Die anschließende Erosion reduziert das Bild so weit, dass es möglichst nahe an das Original herankommt. Die zum Schließen duale morphologische Basis-Operation des Öffnens (*opening*) [469] wurde nach dem Schließen auf die Abb. 3.15 e angewandt. Das Ergebnis ist Abb. 3.15 f zu entnehmen und zeigt bei der Nutzung des gleichen strukturierenden Elementes eines Kreises mit dem Durchmesser von 5 px die Eigenschaft des Schließens als kantenglättende Operation. Als guter Wert für den Durchmesser des strukturierenden Kreises für das Schließen und Öffnen hat sich $1/4$ der Kopflänge bewährt, da auch teilweise erkannte

farbige Kopfbedeckungen nicht ganz eliminiert werden und trotzdem die Glättung so groß ist, dass das Ergebnis nah an die approximierende Ellipse herankommt.

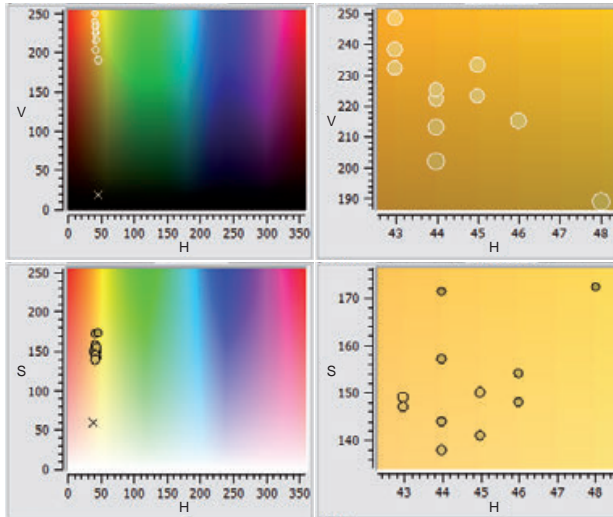


Abb. 3.16: Verteilung der Farben der Kopfbedeckung der in Abb. 3.15 b detektierten Personen im HV- (obere Zeile) und HS-Diagramm (untere Zeile), wie sie in PeTrack angezeigt werden können

Abb. 3.16 zeigt die Verteilung der Farben am Mittelpunkt der approximierenden Ellipsen aller in Abb. 3.15 detektierten Personen. Gut zu erkennen ist, dass sowohl im Farbton-Helligkeits- (obere Zeile) sowie im Farbton-Sättigungs-Diagramm (untere Zeile) die Farbtonvariation von [43,48] sehr gering ist. Die farbigen Kreise in den Diagrammen und in Abb. 3.15 b geben die detektierte Farbe an. Die Variation des Farbtönen verringert sich weiter, wenn durch das anschließende Tracking der Mittelwert der Farbe der Kopfbedeckung entlang des Laufweges ermittelt wird. Für das Beispiel in Abb. 3.15 mit der Farbverteilung von Abb. 3.16 reduziert sich die Anzahl der Farbtöne auf drei im Intervall von [44,46].

Die farbige Kopfbedeckung kann stabil detektiert werden. Jedoch hat der Farbmarker gegenüber den Markern mit Strukturelementen in der jetzigen Implementation zwei Probleme, die durch ein optimiertes Verfahren verringert, aber nicht behoben werden können. Bei hohen Dichten können Kopfbedeckungen insbesondere weiter entfernt von der optischen Achse zusammenlaufen, was bei strenger Parameterwahl zu nicht erkannten Personen führt, da das Seitenverhältnis der approximierenden Ellipse oder die Anzahl der die Kopfbedeckung ausmachenden Pixel zu groß ist. Eine Analyse großer gleicher Farbbereiche zur Erkennung von mehreren Personen innerhalb eines Farbbereiches könnte die Anzahl detektierter Personen erhöhen.

Das zweite Problem stellt der je nach Winkel zur optischen Achse unterschiedliche sichtbare Teil der Kopfbedeckung dar. Dadurch ist die detektierte Kopfposition nicht so exakt und stabil wie bei den komplexeren Markern 1, 2 und 3. Bei den Markern mit Strukturelementen ist das die Position bestimmende Markerelement immer komplett zu sehen (z. B. für Kasernen-Marker (Nr. 1) das Kreuz) und somit befindet sich der Mittelpunkt der approximierenden Ellipse immer an der selben Stelle des Kopfes. Beim Farbmarker hingegen umschreibt die Ellipse je nach Blickwinkel einen anderen Teil der Kopfbedeckung, so dass der Mittelpunkt der Ellipse auf dem Kopf wandert (siehe Abschnitt 3.6.4.2). Eine winkelabhängige Anpassung der angenommenen Position könnte dieses Problem reduzieren, aber nie so stabile Ergebnisse liefern wie strukturierte Marker. Würde die Farbfläche nicht die gesamte Kopfbedeckung ausmachen, könnte der Fehler ebenfalls je nach Größe gemildert werden. Die Erstellung des Markers wird dadurch aber erschwert und der Marker nähert sich mit den Vor- und Nachteilen den Markern mit Strukturelementen an.

3.3 Tracking

Im vorangegangenen Abschnitt wurde vorgestellt, wie die Personendetektion mittels Marker in einem Einzelbild bzw. Frame erfolgt. Für die aufgenommenen Videos bzw. Bildfolgen soll nun erläutert werden, wie erkannte Personen in diesen Sequenzen verfolgt werden, um am Ende für jede durch die Aufnahme laufende Person ihren kompletten Laufweg (Trajektorie) erfasst zu haben (siehe Abb. 3.2 e und Abb. 3.3 e). Diese Standortverfolgung der Personen wird auch Tracking genannt. Dabei wird in diesem Abschnitt das Tracking auf Bildebene beschrieben, später in Abschnitt 3.4 wird unter Zuhilfenahme der in Abschnitt 3.1.1 erläuterten Vermessung die Umrechnung in ein Weltkoordinatensystem angegeben.

Das Tracking erlaubt neben dem Verfolgen von automatisch erkannten Personen auch die Verfolgung von händisch markierten Personen wie es in Versuchsreihe Nr. 5 aus Tab. 4.1 durchgeführt wurde. Dieses halbautomatische Verfahren lässt sich auch für Feldstudien nutzen.

Sei $p'_i(t) \in \mathbb{R}^2$ [px] der Ort auf der Bildebene der Person $i \in \mathbb{N}_N = \{1, 2, 3, \dots, N\}$ zum Zeitpunkt $t \in [t_i^{\text{in}}, t_i^{\text{out}}]$. N ist dabei die Anzahl der im Video verfolgten Personen und $t_i^{\text{in}}, t_i^{\text{out}} \in \{t_j = (j-1)\Delta t \mid j \in \mathbb{N}_F\}$ die Ein- und Austrittszeit der Person i in das bzw. aus dem Sichtfeld der Kamera, wobei Δt der zeitliche Abstand der F Einzelbilder bzw. Frames B_j der zugrundeliegenden Sequenz $\{B_j \mid j \in \mathbb{N}_F\}$ der Gesamtlänge $(F-1)\Delta t$ ist. Die Menge $p' = \{p'_i \mid i \in \mathbb{N}_N\}$ sind alle ermittelten Laufwege.

Ist $p'_i(t_j)$ im Bild B_j bekannt, so wird beim Tracking ein um $\vec{d} \in \mathbb{R}^2$ [px] versetzter Bildpunkt

$$p'_i(t_{j+1}) = p'_i(t_j) + \vec{d} \quad (3.9)$$

im Folgebild B_{j+1} gesucht, dessen Farbwert $B_{j+1}(p'_i(t_{j+1}))$ ähnlich ist. Dabei können je nach Farb- raum und -tiefe des Bildes statt des Farbwertes auch der Hellwert (z. B. bei Grauwert-Bildern) genutzt werden.

Wie die Erkennung geschieht das Tracking auf Subpixel-Ebene durch bilineare Interpolation. Die Pixel-Einheit px ist in dieser Arbeit somit reell, obwohl ein Pixel (*picture element*) die kleinste Bildeinheit beschreibt.

Abb. 3.17 zeigt die Subpixel-Genauigkeit des Trackings für den Translationsvektor $\vec{d} = (d) \in \mathbb{R}$ für den vereinfachten eindimensionalen Fall und damit linearer Interpolation. Für B_j liegt der zu verfolgende Marker (schwarzes Pixel) bei $p'_i(t_j) = 1,5$ px. Für die Pixel des Folgebildes B_{j+1} ergibt sich bei einer Translation um \vec{d} auf $p'_i(t_{j+1}) = p'_i(t_j) + \vec{d}$:

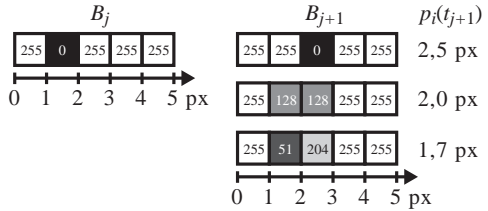


Abb. 3.17: Eindimensionales Tracking des schwarzen Pixels von Bild B_j (Koordinate $p'_i(t_j) = 1,5$ px) nach Folgebild B_{j+1} (Koordinate $p'_i(t_{j+1})$) mit Subpixel-Genauigkeit; der Grauwert $B(u) \in [0, 255]$ ist in die fünf Pixel eingeschrieben

$$B_{j+1}(u) = (1 - (d - \lfloor d \rfloor))B_j(u - \lfloor d \rfloor) + (d - \lfloor d \rfloor)B_j(u - \lfloor d \rfloor - 1). \quad (3.10)$$

Aufgrund von Abbildungsfehlern und zur Erhöhung der Robustheit wird ein Bereich von $2\omega + 1 \in \mathbb{N}$ um diese Punkte auf Ähnlichkeit überprüft, so dass folgender Fehler zu minimieren ist:

$$e_i(t_j) = \operatorname{argmin}_{\vec{d}} \sum_{\omega_x = -\omega}^{\omega} \sum_{\omega_y = -\omega}^{\omega} (B_j(p'_i(t_j) + (\omega_x, \omega_y)^T) - B_{j+1}(p'_i(t_j) + \vec{d} + (\omega_x, \omega_y)^T))^2. \quad (3.11)$$

Eine Rotation des Objektes kann auf diese Weise nicht abgebildet werden und würde zu einem größeren Fehler führen. Wie weiter unten beschrieben wird, ist die Bewegung inklusive Rotation jedoch zwischen zwei Folgebildern gering.

Für die Fehlerminimierung wurde der pyramidenförmige iterative Lucas-Kanade-Merkmal-Verfolger (*pyramidal iterative Lucas Kanade feature tracker*) [448] eingesetzt, da sich dieser als akkurat, robust und schnell (siehe Abschnitt 3.6.8) herausgestellt hat. Andere bekannte Methoden zum Tracking oder zur Verbesserung der Ergebnisse einer Verfolgung sind beispielsweise Kalman-Tracker [452], Mean-Shift-Tracker [454], Camshift-Tracker (*continuously adaptive mean-shift tracker*) [453, 470] oder Tracker speziell für unstrukturierte Bewegungen in Menschenmengen [68].

Das pyramidenförmige iterative Lucas-Kanade-Verfahren sucht dabei in Bereichen gleicher Größe mit Subpixel-Genauigkeit in rekursiven Gauß-Pyramiden [471]

$$B^L(u, v) = \sum_{i=-1}^1 \sum_{j=-1}^1 2^{-(2+|i|+|j|)} B^{L-1}(2u-i, 2v-j) \quad (3.12)$$

mit $B^0 = B$ und $\vec{u}^L = 2^{-L}\vec{u}$, beginnend bei B^{L-1} mit den Pixelkoordinaten $\vec{u}^{L-1} = 2\vec{u}^L$, und propagiert dabei das Ergebnis als Startwert für die nächst höhere Auflösungsebene.

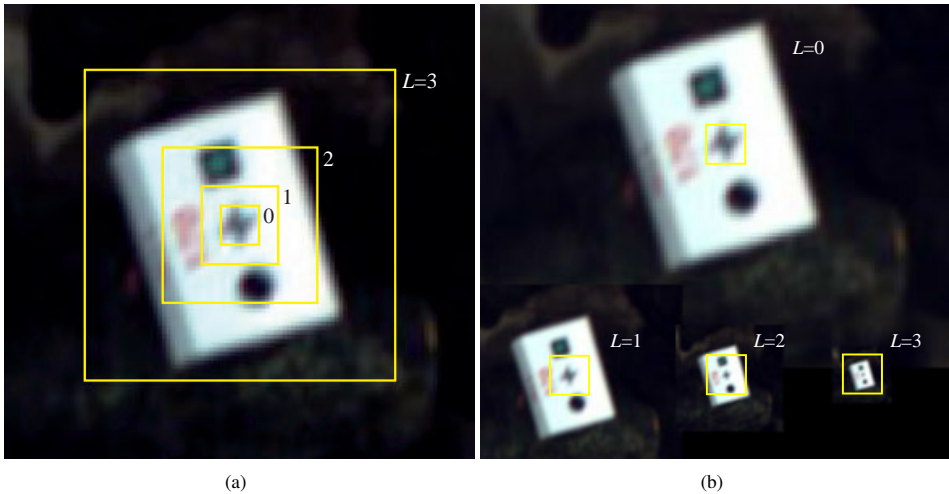


Abb. 3.18: Bildausschnitt mit Vergleichsbereichen in Gauß-Pyramiden am Beispiel des Kasernen-Markers (Nr. 1); (a) zeigt die bei kleiner werdender Ebene L fortschreitende Verringerung des realen Bereichs, der für den Vergleich zweier Folgebilder herangezogen wird, beginnend mit dem gesamten Kopf und Beschränkung auf den zentralen Marker in Ebene $L = 0$; (b) zeigt, dass die Größe des Bereichs zur Fehlerbestimmung gemäß Gleichung 3.11 auf Pixelebene gleich bleibt, aber sich stattdessen die Bildgröße bei steigender Ebene in den Gauß-Pyramiden jeweils halbiert

Die Größe des Vergleichsfensters bzw. der Wert ω wird für die Kopfverfolgung so gewählt, dass auf der niedrigsten Ebene $B^{L_{\max}}$ der ganze Kopf inklusive der Kopfbedeckung und auf der letzten Ebene der Gauß-Pyramide B^0 das Markerelement bei der Fehlerbestimmung gemäß Gleichung 3.11 herangezogen wird (siehe Abb. 3.18). Um bei geneigtem und gedrehtem Kopf die ganze Kopfbedeckung im Vergleichsfenster in $B^{L_{\max}}$ zu haben, wird die Abmessung 50% größer als die durchschnittliche Kopflänge von 21 cm gewählt. Um in der Ebene B^0 beispielsweise genau nur das zentrale Markerelement

von Marker Nr. 1 oder 2 von ca. 4 cm Breite zu vergleichen, sind vier Ebenen ($L_{\max} = 3$) der Gauß-Pyramiden nötig: $1,5 \cdot 21 \text{ cm} / 2^3 \approx 4 \text{ cm}$. Vier Ebenen haben sich für diese Marker zudem als guter Kompromiss zwischen Geschwindigkeit und Robustheit herausgestellt. Ein Bereich kann gut verfolgt werden, wenn dieser eine klare Textur aufweist. Dies ist durch die vom Hintergrund abgesetzte Kopfbedeckung und die Markerelemente gegeben. L_{\max} und ω können in *PeTrack* frei gewählt und die resultierenden Gauß-Pyramiden visualisiert werden.

Das auf dem optischen Fluss aufsetzende verwendete Lucas-Kanade-Verfahren zur Standortverfolgung setzt drei Dinge voraus:

1. gleichbleibende Helligkeit des zu verfolgenden Objektes zwischen aufeinanderfolgenden Bildern,
2. möglichst geringe Objektbewegung oder Nutzung von Pyramiden zur Bewegungsreduzierung auf Pixelebene und
3. räumliche Kohärenz der Pixel (nah beieinander liegende Bildpunkte gehören zum selben Objekt).

Punkt 3 ist bei den zu verfolgenden Köpfen gegeben. Auch Punkt 1 ist zwischen Folgebildern für gewöhnlich erfüllt. Bei nicht gleichmäßiger Ausleuchtung der Beobachtungsfläche kann es jedoch durch harte Schatten, wie in Abb. 3.19 zu sehen, zum Aufhängen der Verfolgung an der Schattenkante kommen, da diese das Helligkeitsprofil der Umgebung dominiert. Auch ist in ähnlichen Situationen ein Abbruch des Trackings möglich, da der Fehler gemäß Gleichung 3.11 zu groß wird. Würden die zur Erkennung der Marker herangezogenen Isowerte der Helligkeit klein genug gewählt werden, würde für den abgebildeten Fall der Erkennungsschritt das korrekte Verfolgen der Person ermöglichen (siehe Marker am rechten Rand von Abb. 3.12).

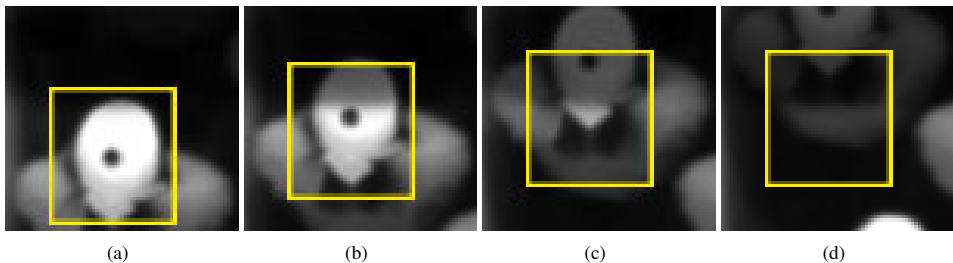


Abb. 3.19: Scheiterndes Tracking einer Person, die in einen starken Schatten eintritt

Auch kann das Tracken sich bei temporärer Teilverdeckung an den im Weg befindlichen Objekten festsetzen und damit die zu verfolgende Person verlieren. Die Teilverdeckung kann z. B. durch das Unterwandern von überkopf angebrachten Objekten wie Kabeln oder Traversen, durch Aufbauten wie Wände oder gegenseitig durch unterschiedlich große Personen entstehen. Eine Mindestquote der Erkennung einer Person entlang seines Laufweges oder das Eliminieren von Trajektorien, die einen vorgegebenen eigentlich zu durchlaufenden Bildbereich nicht komplett durchlaufen, kann dieses Problem beheben und ist in *PeTrack* beim Test der Laufwege oder beim Export implementiert (siehe Abschnitt 3.3.4).

Punkt 2 ist aufgrund der Bildwiederholrate von minimal 15 f/s (siehe Bereich Bildwiederholrate im Abschnitt 4.1.1.5) und den dabei maximal zurückgelegten 0,1 m/f bei einer freien Gehgeschwindigkeit von zügigen 1,5 m/s ebenfalls gegeben. Die Kopfschwankungen sind für die ausschlaggebende Maximalgeschwindigkeit irrelevant und zudem bei geringen Geschwindigkeiten größer als bei hohen Geschwindigkeiten. Die in Folgebildern zueinander nächsten Marker identifizieren somit die gleiche Person. Abb. 3.20 c zeigt das Differenzbild der aufeinanderfolgenden Bilder a und b bei Experimenten der Versuchsreihe Nr. 3 mit Marker Nr. 1. Das Differenzbild stellt die geringen Unterschiede zwischen zwei Folgebildern dar, auch wenn es sich in diesem Fall um eine Engstelle handelt und die Geschwindigkeit nicht maximal ist, so sind bei den verwendeten Kameras die maximalen Positionsunterschiede korrespondierender Pixel nur 15 px/f, wobei der gesamte Marker 35 px \times 53 px misst (siehe Video Nr. 14).

Das Differenzbild könnte den Anschein erwecken, dass es sich zur Lokalisierung von Personen ggf. sogar ohne Marker eignet. Dies ist leider aus dem gleichen Grund nicht möglich wie die direkte Nutzung des optischen Flusses zur Personendetektion (siehe Abschnitt 2), da insbesondere Laufwege bei hohen Dichten und somit auch Stausituationen bestimmt werden sollen und damit eine verschwindende Pixeldifferenz über längere Zeit, wie sie z. B. bei der grün umrandeten Person in Abb. 3.20 c zu erkennen ist, existieren kann.

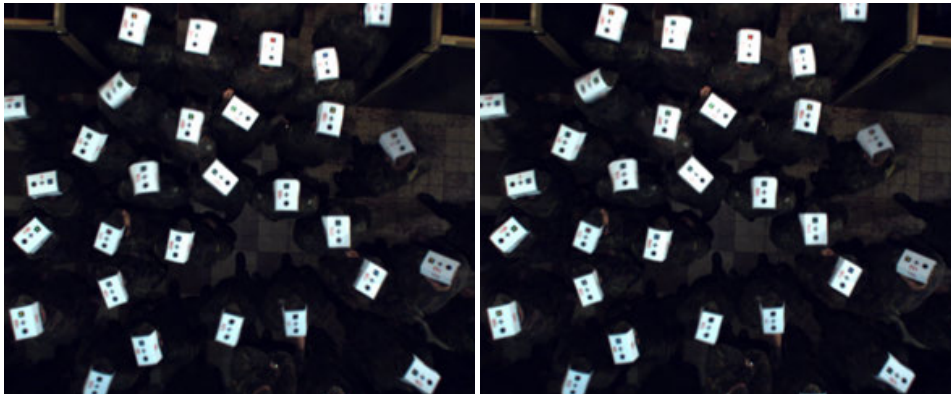
Das Tracking ist in PeTrack so implementiert, dass ein Verfolgen auch beim Spulen und dabei Überspringen von Bildern in einer Bildfolge funktioniert, um das Erzeugen der Laufwege zu beschleunigen. Für ein stabiles Tracking muss dabei der nächstgelegene Marker im Folgebild die gleiche Person identifizieren. Bei hohen Dichten bewegen sich die Köpfe wenig, bei geringen Dichten wird die freie Gehgeschwindigkeit erreicht. Bei einem an der Kapazitätsgrenze liegenden, spezifischen Fluss passieren am meisten Personen pro Zeit eine bestimmte Stelle und nähern sich somit am schnellsten der Position des Vorausgehenden. Daher ist für die Bestimmung der auslassbaren Bilder beim Tracking der maximale Personenfluss von ca. 1,5/(m s) bei einer Geschwindigkeit von $v = 0,7$ m/s und einer Personendichte von $\rho = 2/\text{m}^2$ [472] heranzuziehen. Bei einem Einzelbildabstand im Bereich von $\Delta t \in [1/15 \text{ s}, 1/30 \text{ s}]$ und gemäß Gleichung 4.1 einem Personenabstand von $d_p = 1/\sqrt{\rho} \approx 0,7$ m dürfen somit

$$\left\lfloor \frac{d_p}{2\Delta t v} \right\rfloor - 1 \in [6, 14] \quad (3.13)$$

Bilder übersprungen werden. PeTrack beschränkt dies auf 4 übersprungene Bilder und interpoliert die Position des Laufweges in den ausgelassenen Bildern linear. Die Begrenzung ermöglicht noch, die Schwankbewegung des Kopfes mit einer Schrittdauer von im Schnitt etwas unter einer Sekunde [156] abzubilden. Die Beschleunigung der Laufwegebestimmung durch das Überspringen von Einzelbildern wurde bisher selten durchgeführt, da die Laufwege im Nachgang der Erfassung aus den Aufnahmen gewonnen werden und damit der Prozess nicht so zeitkritisch ist, dass auf die höhere Genauigkeit durch Berücksichtigung aller Einzelbilder verzichtet werden muss.

Da jedoch das Tracking bei kontrastreichen Markern sehr gute Ergebnisse liefert und das Tracking schneller abläuft als die Erkennung, ist es nahezu ohne Genauigkeitsverlust für die Laufwege möglich, den Erkennungsschritt nur im Abstand einiger Einzelbilder durchzuführen und dazwischen die Person ausschließlich zu verfolgen. Der Einzelbildabstand kann in PeTrack eingestellt werden und ist standardmäßig 5 Bilder. Abb. 3.21 zeigt den ermittelten Laufweg einer Person in rot, wenn die Pixelposition der Person durch die Markererkennung erfolgt, und in grün, wenn die Person in der Bildmitte einmalig detektiert und ansonsten nur getrackt wird.

Dabei geschieht die Verfolgung immer ausschließlich zwischen zwei Bildern und bezieht kein Wissen über das Aussehen der Marker an anderer Stelle des Laufweges insbesondere zum Zeitpunkt der



(a)

(b)



(c)

Abb. 3.20: Differenzbild (c) der Bilder (a) und (b) zur Visualisierung der Bewegung der zu verfolgenden Marker zwischen aufeinanderfolgenden Einzelbildern einer Sequenz; das grün umrandete Rechteck zeigt eine sich kaum bewegende Person mit nahezu verschwindender Pixeldifferenz

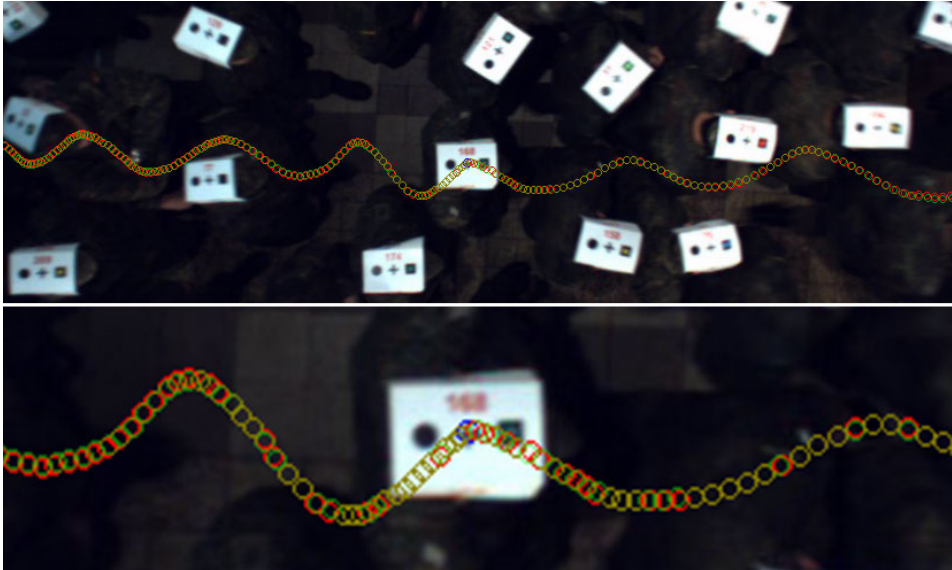


Abb. 3.21: Vergleich der ermittelten Laufwege: Markererkennung in jedem Einzelbild (Pfad roter Kreise) und Erkennung in Bildmitte eines einzelnen Bildes und anschließendem reinen Tracking (Pfad grüner Kreise); Mischton, wenn Pfade aufeinander liegen

Erkennung mit ein. Dies hat den Vorteil, dass leichte Änderungen im Aussehen der Kopfbedeckung (z. B. durch Licht, Kopfnähe) kontinuierlich berücksichtigt werden können.

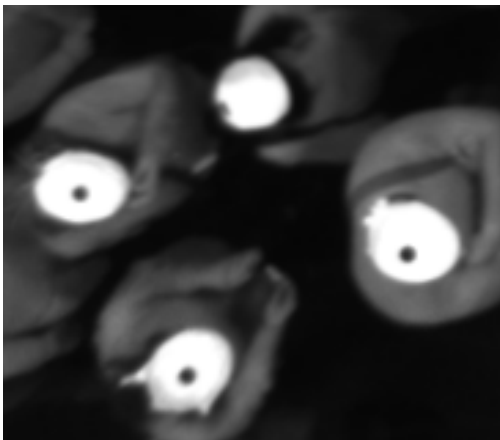


Abb. 3.22: Bei der obersten Person könnte beim reinen Tracking die zur Optimierung dienende Positionsverschiebung in Richtung des dunkelsten Bildpunktes des mittleren Markerelementes dazu führen, dass die Trajektorie auf den dunklen Hintergrund abdriftet

Um ein mögliches Abdriften beim reinen Tracking von einem zentralen dunklen Markerelement (vorhanden in Marker Nr. 1 und 2) zu verhindern (z. B. bei starkem Schatten wie in Abb. 3.19), kann in *PeTrack* eingestellt werden, dass im zentralen Bereich des mittigen Markerelementes die verfolgte Position in Richtung des dunkelsten Bildpunktes verschoben wird. Beim Hermes-Marker (Nr. 2) muss bei dieser Optimierung darauf geachtet werden, dass alle Personen die Kopfbedeckung so tragen, dass im gesamten Bild um das zentrale Markerelement die restliche weiße Kopfbedeckung zu sehen ist, da ansonsten diese Optimierung dazu führen kann, dass das Tracking auf den noch dunkleren Hintergrund kontinuierlich verschoben wird. Kritisch könnte die angestrebte Verbesserung somit für die obere Person in Abb. 3.22 sein.

3.3.1 Neuberechnung und Qualität

In einigen Fällen kommt es vor, dass die Position einer Person wiederholt im gleichen Bild bestimmt wird. Um in diesen Fällen entscheiden zu können, welche Lokalisierung gewählt wird, wird jeder ermittelten Position jeder Person $i \in \mathbb{N}_N$ in jedem Bild $B_j, j \in \mathbb{N}_F$ eine Qualität $q_i(t_j)$ zugewiesen. Der Wertebereich der Qualität liegt im Intervall von $[0, 100]$, wobei 100 der höchsten Qualität entspricht, und wie folgt vergeben:

$q = 100$: wenn eine Person mit allen Markerelementen erkannt wurde

oder

wenn in der grafischen Oberfläche von PeTrack die Position von Hand festgelegt wird,

$20 \leq q \leq 80$: wenn eine Person getrackt wurde; dabei wird die Qualität von 80 vergeben, wenn kein Fehler gemäß Gleichung 3.11 vorliegt, und bis zum Wert von 20 wie folgt verringert; der flächenbereinigte Fehler für Grauwertbilder mit $B(u, v) \in [0, 255]$ ist

$$e'_i(t_j) = \sqrt{\frac{e_i(t_j)}{(2\omega + 1)^2}} = (2\omega + 1)^{-1} \sqrt{e_i(t_j)} \quad (3.14)$$

und aufgrund der Subadditivität der euklidischen Norm gilt $e'_i(t_j) \in [0, 255]$. Als Qualität wird nun

$$q_i(t_j) = \begin{cases} 80 - 2e'_i(t_j), & \text{falls } e'_i(t_j) < 30 \\ 20, & \text{sonst} \end{cases} \quad (3.15)$$

angenommen,

$q = 0$: wenn eine Person anhand der Kopfbedeckung, aber ohne Detektion der markerspezifischen Elemente erkannt wurde; diese Erkennungsstrategie kann in PeTrack deaktiviert werden

oder

wenn die Position durch Extrapolation aufgrund einer unmöglichen Trackingposition bestimmt wurde.

Die erneute Positionsbestimmung einer Person überschreibt die alte Position, wenn die Qualität besser ist oder 100 beträgt. Bei der Qualität von 100 wird somit immer überschrieben, damit manuell gesetzte oder durch Marker erkannte Positionen von Hand verändert werden können. Die erneute Bestimmung kann in folgenden Fällen auftreten:

- für einen durchgängigen Laufweg muss die Person durch die gesamte Sequenz getrackt werden; findet eine Lokalisierung durch die Markererkennung ($q = 100$) statt, so überschreibt die lokalisierte Position die durch Tracking ($20 \leq q \leq 80$) ermittelte Position,
- in PeTrack kann ein Qualitätswert festgelegt werden, bis zu dem bei einem wiederholten Durchlaufen der Sequenz zur Laufwegebestimmung die Person erneut getrackt und bei erhöhter Qualität die alte Position überschrieben wird; dies könnte z. B. auftreten, wenn bei einer existierenden Trajektorie eine Position händisch verändert wird und ab diesem Frame das Tracking neu angestoßen wird, oder wenn beim Rückwärtsdurchlaufen der gesamten Sequenz zur Vervollständigung der Laufwege insbesondere zum Rand hin und vor der ersten Erkennung einer Person das Tracking von anderen detektierten Positionen aus wie in Abb. 3.23 skizziert durchgeführt wird,



Abb. 3.23: Laufwegbestimmung beim Vorwärts- und anschließendem Rückwärtsdurchlaufen einer Sequenz; beim späteren wiederholten Verfolgen gemäß der unteren Pfeile könnte für eine höhere Qualität (rote Zahl) die alte durch die neue Position ersetzt werden

- die Position einer bereits lokalisierten Person ($q \leq 100$) wird von Hand neu festgelegt ($q = 100$); dies überschreibt immer die zuvor bestimmte Position,
- wird eine Sequenz z. B. stückweise verarbeitet oder geht eine Person beim Tracking verloren, wird aber beim Rückwärtsdurchlaufen wiedergefunden, so kann es vorkommen, dass zwei Laufwege verschmolzen (*merge*) werden müssen, da auch hier die Position einer Person in einem Bild doppelt bestimmt wurde; liegen zu mehreren Zeitpunkten für die zu verschmelzenden Laufwege Positionen vor, so wird entweder die längere Trajektorie übernommen oder bei sich zeitlich einseitig überlappenden Laufwegen, wie beim Verschmelzen zweier Kameraperspektiven, eine linear gewichtete neue Position bestimmt (siehe Abschnitt 3.3.5); das Verschmelzen kann deaktiviert werden, um zu verhindern, dass Personen durch die perspektivische Sicht zum Rand für unterschiedlich große Personen oder an Treppen hintereinander verschwinden und damit die bestimmte Position beider Personen einer zugeschrieben würde.

Eine Position wird als identische Person angenommen, wenn die neue Position im Kopfbereich der zuvor erkannten Person liegt. Als Kopfbereich wird ein Kreis vom Durchmesser der durchschnittlichen Kopflänge von ca. 21 cm gewählt. Wie viele Pixel die Kopfgröße ausmachen, wird mit Wissen des Kopfabstandes zur Kamera bzw. der Körpergröße entsprechend der inversen Betrachtung von Abschnitt 3.4 bestimmt.

Für gewöhnlich werden alle in einem Bild bis zu diesem Zeitpunkt erkannten Personen zum nächsten Bild hin verfolgt. Gibt es bereits für das Folgebild erkannte Personen (Trajektorien mit Positionen in der Zukunft), so kann in *PeTrack* das Tracking deaktiviert werden bzw. eine obere Qualitätsschranke (standardmäßig $q = 50$) angegeben werden, bis zu der eine wiederholte Verfolgung durchgeführt werden soll. Zudem kann in *PeTrack* das Tracking auch nur für einzelne Personen veranlasst werden.

Die Verfolgung bricht ab, wenn die Person aus dem Bild herausläuft oder einen Maximalfehler von $e'_i(t_j) > 40$ für die durchschnittliche Differenz der Pixelfarbe überschreitet. Die Verfolgung wird über den für die Erkennung eingeführten relevanten Bildbereich hinaus durchgeführt. Der maximale Fehler als Abbruchkriterium müsste für kontrastarme Bilder und damit Abbildungen der Marker verringert werden, da dort die Differenz der Pixelfarbe immer gering ausfällt. Dies ist nicht implementiert, da bei allen Aufnahmen die Belichtung so eingestellt wurde, dass die Marker kontrastreich gegenüber dem Hintergrund erscheinen.

3.3.2 Extrapolation

Das Tracking geschieht wie zuvor erwähnt nur zwischen zwei Bildern ohne Vorwissen über den bisherigen Laufweg (im Gegensatz zu einem Kalman-Tracker [452]). Sollte sich dabei herausstellen, dass die Geschwindigkeit sprunghaft ansteigt (z. B. ein Überspringen auf eine andere Person), muss dies daher separat abgefangen werden. Eine sprunghafte Verringerung tritt nicht auf, da eine ähnliche Struktur näher als der zu verfolgende Kopf nicht vorkommt. Bei einem sprunghaften Anstieg wird statt der ermittelten Tracking-Position eine Extrapolation des Laufweges durchgeführt. Konkret wird

ab einer spontanen Verdopplung der Geschwindigkeit von einem auf den anderen Tracking-Schritt die neue Tracking-Position verworfen und die Geschwindigkeit inklusive der Richtung als gleichförmig angenommen:

$$p'_i(t_{j+1}) = p'_i(t_j) + (p'_i(t_j) - p'_i(t_{j-1})) = 2p'_i(t_j) - p'_i(t_{j-1}) \quad (3.16)$$

Da diese Annahme nur kurzfristig der eigentlichen Bewegung nahe kommt, ist die Extrapolation in *PeTrack* nur für zwei aufeinanderfolgende Bilder erlaubt. Danach bricht die Verfolgung ab, wenn die Position nicht erneut detektiert wurde. Zudem wird der Test der abrupten Geschwindigkeitssteigerung nur ab einer Mindestgeschwindigkeit durchgeführt, da die Verdopplung der Geschwindigkeit nahe des Stillstandes und natürlich beim Stillstand selber eintreten kann bzw. muss. Nach der Extrapolation wird die oben beschriebene Optimierung der Position durch die Verschiebung in Richtung des nächst gelegenen dunklen Pixel durchgeführt.

3.3.3 Fehlbilder

Wird die Bildfolge nicht direkt an der Kamera durch ein geschlossenes System aufgezeichnet, sondern von einer Industriekamera, die über einen Bus das Signal auf einem Rechner speichern muss, kann es bei nicht adäquater Auslegung der Einzelkomponenten passieren, dass nicht alle von der Kamera angeforderten Bilder aufgezeichnet werden können (siehe Abschnitt 4.1.1.9). Dies ist trotz ausgiebiger Tests im Vorfeld bei den Kasernen-Versuchen (Nr. 3) vorgekommen. Die Anzahl aufeinanderfolgender fehlender Bilder war dabei nie größer als 2 und in der Summe nie mehr als 4 innerhalb eines Versuchs (bei durchschnittlicher Länge der Bildfolgen von 2500 Frames).

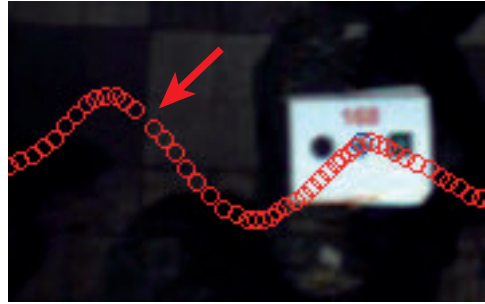


Abb. 3.24: Lücke in der Trajektorie aufgrund fehlendem Bild in der Aufzeichnung, gekennzeichnet durch einen roten Pfeil; die roten Kreise zeigen die Position der Person in allen Bildern der Aufzeichnung

Abb. 3.24 zeigt das Resultat in der Trajektorie für ein fehlendes Einzelbild. In der abgebildeten Trajektorie zeigen die roten Kreise den Verlauf der Position der Person für alle vorangegangenen und kommenden Bilder an. Der rote Pfeil zeigt auf einen Sprung in der Trajektorie aufgrund eines nicht aufgezeichneten Bildes.

Ist nicht wie bei den Hermes-Versuchen (Nr. 6) der Zeitstempel jedes Bildes mitgeschrieben worden, können Fehlbilder und deren Anzahl durch Vergleich der Abstände aufeinanderfolgender Punkte innerhalb aller Trajektorien detektiert werden. Für einen Sprung zwischen dem Bild zum Zeitpunkt t und dem Folgebild zum Zeitpunkt $t + \Delta t$ wird der Abstand zu vorangegangenen $\|p'_i(t) - p'_i(t - \Delta t)\|$ und nachfolgenden $\|p'_i(t + 2\Delta t) - p'_i(t + \Delta t)\|$ Punkten in den Trajektorien mit der möglichen Lücke $\|p'_i(t + \Delta t) - p'_i(t)\|$ für alle N Personen $i \in \mathbb{N}_N$ verglichen, die zwischen t und $t + \Delta t$ eine Trajektorie besitzen. Die gerundete Anzahl $g(t)$ ausgelassener Bilder zwischen t und $t + \Delta t$ kann damit wie folgt ermittelt werden:

$$g(t) = \left\lfloor \frac{2}{N} \sum_{i=1}^N \frac{\|p'_i(t + \Delta t) - p'_i(t)\|}{\|p'_i(t) - p'_i(t - \Delta t)\| + \|p'_i(t + 2\Delta t) - p'_i(t + \Delta t)\|} - 0,5 \right\rfloor. \quad (3.17)$$

Der angenommene Schwellenwert ist robust und führt zu keinen Fehldetektionen, da bei allen Experimenten außer für die ersten und letzten Bilder die Anzahl gleichzeitig verfolgter Personen im Bild

groß genug ist. Fehlstellen in den Trajektorien können linear interpoliert werden, weil nie mehr als zwei Fehlbilder detektiert wurden und in der betroffenen Versuchsreihe Nr. 3 die Bildwiederholrate mit 25 ausreichend groß war.

3.3.4 Bereinigung

Neben der Möglichkeit, Fehlstellen in Trajektorien automatisch einfügen zu lassen, werden viele weitere Optionen zur Behandlung der Laufwege in *PeTrack* angeboten. In Anhang D.3 wird die Nutzung der Software detaillierter vorgestellt. Abb. D.5 c zeigt den Reiter zur Steuerung der Parameter für das Tracking und damit die Möglichkeiten, beim Exportieren Bedingungen an die herausgeschriebenen Laufwege zu stellen, Laufwege auf besondere Konstellationen hin zu testen und zur Unterstützung der manuellen Korrektur ausgewählte Daten der Laufwege zu visualisieren. Warum eine manuelle Korrektur implementiert wurde, ist Abschnitt 4.2 zu entnehmen.

Die Korrektur der ermittelten Laufwege kann unterschieden werden in automatische und bei der manuellen Durchsicht unterstützende Verfahren. Die automatischen Verfahren lassen sich noch einmal unterteilen in Methoden, die direkt während des Trackings herangezogen werden, und Methoden, die beim Exportieren der Laufwege auf diese angewendet werden.

Unterstützende Verfahren zur manuellen Korrektur erlauben in *PeTrack* auffällige Stellen zeitlich und räumlich aufzuzeigen. Dabei wird die Bildfolge an die mögliche Fehlstelle bewegt und die auffällige Trajektorie von Person i hervorgehoben. Folgende Untersuchungen sind implementiert worden und zeigen eine Stelle an, wenn:

- die Trajektorie ein Zeitintervall von weniger als einer halben Sekunde umfasst:

$$t_i^{out} - t_i^{in} < s/2,$$

- der Beginn oder das Ende der Trajektorie innerhalb des als relevant gekennzeichneten Bildbereichs liegt (siehe Seite 61); dieser Test kann z. B. für die Versuche 1c und 5c in Abb. 3.8 bzw. 3.9 Hinweise auf Fehler geben, hingegen nicht für 2c, 3c und 4c, da entweder der zentrale Versuchsbereich unterirdisch verlassen wird (2c), nur in den oder aus dem Bereich gegangen wird (3c) oder die Probanden sich nur innerhalb des relevanten Bereichs bewegen (4c),

- Punkte zweier Trajektorien von Person i und j zu einem Zeitpunkt t näher als eine halbe Kopflänge beieinander liegen:

$$\|p_i(t) - p_j(t)\| < 21 \text{ cm}/2;$$

dies kann nur auftreten, wenn in *PeTrack* für nahe Trajektorien kein Verschmelzen (*merge*) aktiviert wurde; die Deaktivierung kann für stark weitwinklige Aufnahmen oder Aufnahmen an Treppen sinnvoll sein, da es für diese Fälle vorkommen kann, dass kleine Personen hinter großen Personen verschwinden (siehe Abschnitt 4.1.1.1),

- ab einer Mindestgeschwindigkeit die Geschwindigkeit zwischen zwei Frames mit einem Einzelbildabstand von Δt doppelt so groß wie in dem davor und dahinter liegenden Zeitintervall ist:

$$\|p'_i(t + \Delta t) - p'_i(t)\| > \|p'_i(t) - p'_i(t - \Delta t)\| + \|p'_i(t + 2\Delta t) - p'_i(t + \Delta t)\|;$$

dieser Test ist nur für Laufwege von markierten Personen ein Hinweis auf mögliche Unregelmäßigkeiten in der Trajektorie; bei unmarkierten Personen, die mittels des Verfahrens in Abschnitt 3.5 detektiert werden, besitzt der Laufweg Sprünge aufgrund schwankender Lokalisation des Kopfes.

Automatische Verfahren, die während des Exportierens der Laufwege herangezogen werden können, sind:

- Suche nach und Einfügen von fehlenden Positionen in den Trajektorien aufgrund von Fehlbildern wegen fehlerhafter Aufnahmen (siehe Abschnitt 3.3.3),
- Ermittlung des Medians der über den Laufweg hinweg ermittelten Personen-Größen statt Nutzung eines während des Trackings angepassten Mittelwertes der Personen-Größe,
- Einschränkung der Personen-Größe auf einen Durchschnittswert für ebene Versuche,
- Löschen von randständigen Punkten einer Trajektorie oder ganzer Trajektorien, wenn für einzelne Punkte oder den ganzen Laufweg keine Personen-Größe ermittelt werden konnte,
- Glättung des Personenabstandes zur Kamera, indem Sprünge mit einer Geschwindigkeit von mehr als $6,4 \text{ m/s}$ (40 cm zwischen zwei Frames bei einer Bildwiederholrate von 16 f/s) in der Höhe durch die nächstgelegene gültige Höhe entlang der Trajektorie oder dem Mittelwert umliegender Abstandswerte ersetzt wird; dies tritt nur bei der Abstandsbestimmung über die Disparität von Stereokameras auf (siehe Abschnitt 3.4.2) und kann nicht von einer natürlichen Personenbewegung herrühren; die Geschwindigkeit in senkrechter Richtung ist z. B. beim Aufstehprozess ca. $0,8 \text{ m/s}$ oder beim Treppensteigen in normaler Geschwindigkeit ca. $0,5 \text{ m/s}$ (siehe Abb. 3.25).

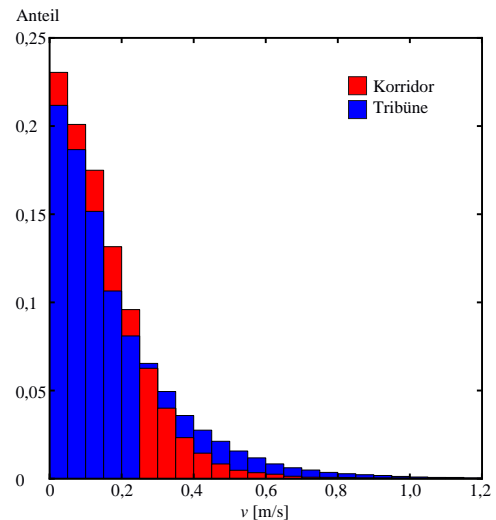


Abb. 3.25: Verteilung der Geschwindigkeit entlang der optischen Achse; beim Experiment entlang eines Korridors rührt die vertikale Bewegung von Größenänderungen während eines Schrittes her; in der Tribüne werden größere Geschwindigkeiten durch das Treppensteigen und den Aufstehprozess erzielt (Versuch gemäß Abb. 3.3)

Während des Trackings sind neben dem bereits beschriebenen Verschmelzen (siehe Abschnitt 3.3.1) oder der Extrapolation (siehe Abschnitt 3.3.2) für Sprünge während des Trackings noch zwei weitere Tests zur Selektion falscher Trajektorien implementiert worden. Führt das Tracking den Laufweg in einen ermittelten Hintergrund (siehe Seite 61 und Abschnitt 3.5.1), so wird die Trajektorie gelöscht, wenn sie mehr als zwei aufeinanderfolgende Frames im Hintergrund verläuft. Für den Fall, dass die Laufwegebestimmung ohne Marker durchgeführt wird, wird eine Trajektorie bereits während des Trackings verworfen, wenn mehr als 80% des Laufweges eine erkannte Person nur verfolgt und nicht wieder erkannt wird.

3.3.5 Verschmelzung

Sollen Versuche erfasst werden, die aufgrund der Größe der Versuchsfläche, der geringen Deckenhöhe, der Verdeckung der Personen oder der zur Personenerkennung benötigten Pixelabdeckung den Einsatz mehrerer Kameras verlangen, so ist eine Überlappung vorzusehen, um zuverlässig alle Personen beim Wechsel von einer zur anderen Kamerasicht verfolgen zu können (siehe Abschnitt 4.1.1.2).

Die Güte der Trajektorien ist bei allen bisher durchgeführten Versuchen auch zum Rand hin so gut, dass eine zuverlässige Zuordnung der Trajektorienpaare möglich ist (siehe Abb. 3.26). Dabei werden die Trajektorienpaare so gewählt, dass der Gesamtfehler des Quadratabstandes aller einander zugehörigen Punkte minimiert wird.

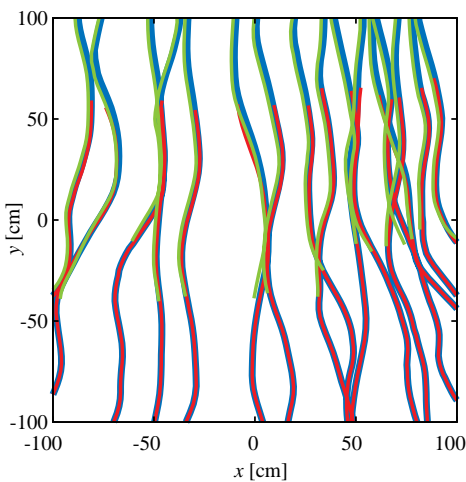


Abb. 3.26: Ausschnitt von Trajektorien zweier überlappender Aufnahmen (rot und grün) aus Versuchsreihe Nr. 3 und die resultierende zusammengeführte Trajektorie (blau), die hinter der roten Trajektorie weiterläuft

Falls die Kameras, aus denen die zu verschmelzenden Trajektorien extrahiert wurden, nicht synchronisiert waren, ist ggf. eine zeitliche Verschiebung zur Zuordnung der Einzelbilder nötig. Wenn die Versuchsreihe komplett aufgezeichnet wurde, kann dies manuell geschehen, indem einmalig kurzzeitige visuelle oder akustische Signale aufeinander gelegt werden (z. B. Versuchsreihe Nr. 7). Bei Aufzeichnungen mit Industriekameras wurde jeder Versuch separat gesichert, so dass die zeitliche Verschiebung automatisch geschehen sollte (z. B. Versuchsreihe Nr. 3). Hierzu wird die Suche nach der kleinsten Fehlerquadratsumme wiederholt durchgeführt, während die Datensätze bildweise innerhalb eines Zeitintervalls gegeneinander verschoben werden. Dabei ist von Vorteil, dass die Aufnahmestarts der Kameras für gewöhnlich zeitlich nur kurz auseinander liegen. Die Verschiebung mit dem kleinsten Fehler wird genutzt. Da die Kameras nicht synchronisiert waren, kann es zu einem maximalen Fehler im zeitlichen Ver-

satz von der Hälfte des Kehrwertes der Bildwiederholrate kommen. Für die Kasernen-Versuche (Nr. 3) ist dies $1/(2 \cdot 25)$ s. Bei einer Normalgeschwindigkeit von 1 m/s würde sich für die Position somit ein Fehler von maximal 2 cm ergeben. Bei gradlinigen unidirektionalen Bewegungen kann dies zum Teil durch eine automatische Verschiebung kompensiert werden, die im folgenden Absatz erläutert wird. Bei den Hermes-Versuchen (Nr. 6) liefen die Kameras synchron und zeichneten zudem die absolute Zeit jedes aufgezeichneten Bildes auf, so dass eine eindeutige zeitliche Zuordnung beim Verschmelzen der Trajektorien gegeben ist.

Nach dem zeitlichen Versatz wird eine den durchschnittlichen Fehler minimierende Transformation eines Trajektorien-Datensatzes gesucht, die entsprechend die extrinsischen Parameter anpasst. Die verwendete Methode zur Suche des optimalen Translationsvektors und der optimalen Rotationsmatrix zur Minimierung der Fehlerquadratsumme aller zueinander gehörenden Punktepaare basiert auf [473]. Dass eine Transformation das Ergebnis verbessert, liegt vornehmlich daran, dass ggf. das manuelle Festsetzen des Koordinatensystems nicht exakt ist, das Ausmessen der Kamerahöhe mit einem Fehler behaftet ist oder auch das senkrechte Ausrichten der Kameras nicht fehlerfrei durchgeführt

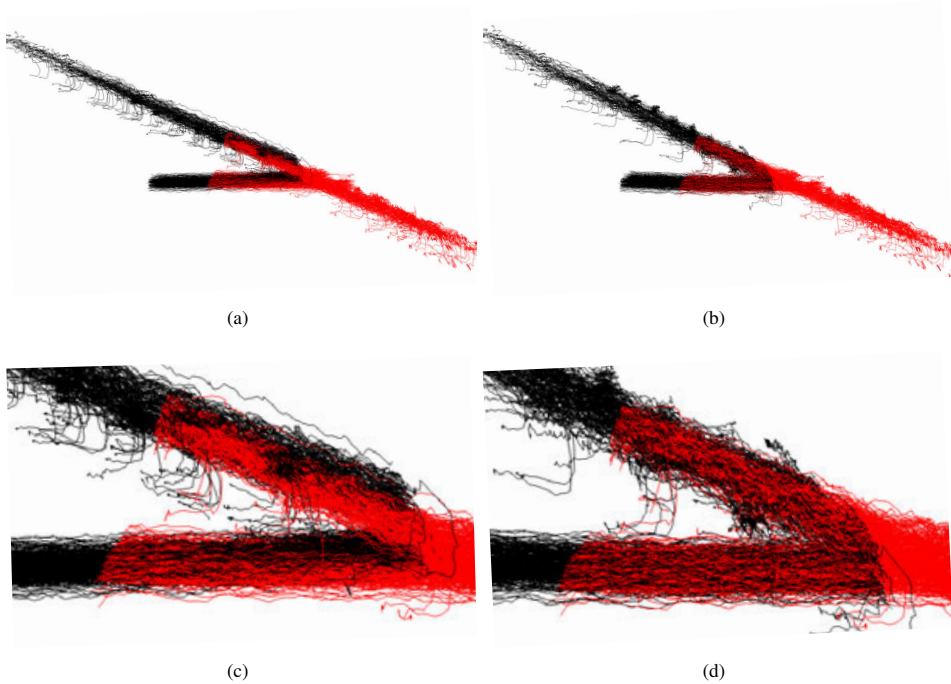


Abb. 3.27: Seitenansicht zweier überlappender Trajektoriensätze eines Experiments beim Verlassen einer Tribüne durch ein Mundloch; die schwarzen Trajektorien zeigen den Treppenaufgang und den kompletten Austritt aus dem Mundloch, die roten Trajektorien vornehmlich den Treppenaufgang; die nach unten herausragenden Teile der Trajektorien zeigen den Beginn der Trajektorien beim Aufstehvorgang der Probanden; (b, d) zeigt die Rotation des schwarzen Datensatzes im Vergleich zu (a, c) zur Minimierung des Fehlers beim Zusammenführen der Trajektoriensätze; (c, d) sind Ausschnitte aus (a, b); vgl. Abb. 4.8

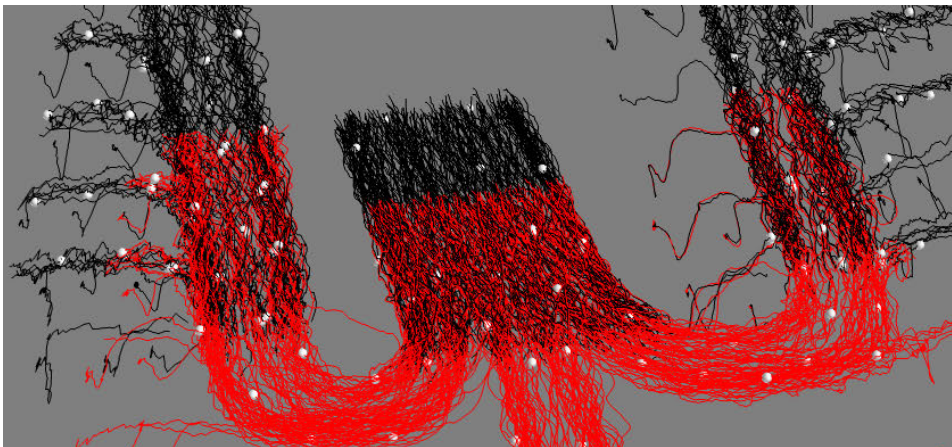


Abb. 3.28: Gedrehte Ansicht von Abb. 3.27 b als Ausschnitt von Abb. 3.71; die weißen Kugeln zeigen die Personenverteilung zum Zeitpunkt der Aufnahme in Abb. 3.70; siehe Video Nr. 45 für den gesamten Versuchsdurchlauf

wurde. Abb. 3.27 zeigt Seitenansichten zweier überlappender Trajektoriensätze eines Experiments an Treppen beim Verlassen einer Tribüne vor und nach der Transformation eines Datensatzes zur Minimierung des Fehlers beim Zusammenführen.

Für alle ebenen Versuche mit überlappenden Kameraperspektiven (Nr. 3 und 6) ergibt sich nach der Transformation ein durchschnittlicher Fehlerabstand zwischen den korrespondierenden Trajektorien von $1 \text{ cm} \pm 0,4 \text{ cm}$ und ein maximaler Fehler von $5 \text{ cm} \pm 2,6 \text{ cm}$ in der 2D-Bewegungsebene. Für alle Experimente an Treppen ist im Raum der durchschnittliche Fehlerabstand $3 \text{ cm} \pm 0,5 \text{ cm}$ und der maximale Fehler $8 \text{ cm} \pm 2,6 \text{ cm}$.

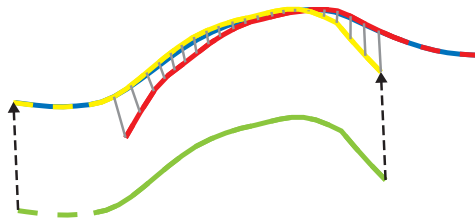


Abb. 3.29: Skizze der Interpolation bei der Verschmelzung von Trajektorien: nach der Verschiebung der grünen Trajektorie entlang des gestrichelten Pfeils findet für die Berechnung der resultierenden blauen Trajektorie eine Interpolation zwischen der roten und gelben Trajektorie statt, wobei das Gewicht zum Ende der jeweiligen Trajektorie abnimmt; die grauen Linien verbinden die zueinander gehörigen Punkte

Der Fehler steigt zum Rand des Bildes aufgrund des größeren Einflusses der Personengröße und einer nie vollkommenen Entzerrung. Um diesen Fehler zu verringern, wird bei dem Verschmelzen der korrespondierenden Trajektorienpaare eine Interpolation durchgeführt, bei der die Trajektorien zum Bildrand weniger Gewicht erhalten. Abb. 3.29 verdeutlicht dies. Alle später im Abschnitt 4.4 verwendeten Daten sind das Resultat von aus mehreren Kameraansichten zusammengeführten Trajektorien.

Eine Beschreibung der Nutzung des Programms zur Trajektorienverschmelzung ist in Anhang D.4 zu finden.

3.4 Positionsbestimmung

In diesem Abschnitt soll die Pixelposition des Kopfes im Bild in eine reale Position im Weltkoordinatensystem überführt werden. Abb. 3.30 zeigt die Überführung anhand der beiden aus PeTrack exportierbaren Datenstrukturen der Trajektorien. Links stehen die direkt aus dem Bild extrahierten Daten wie u. a. die Pixelposition, die Qualität des Trackings, die Position des Farbmarkerelementes und dessen Farbe, also die bisher in diesem Kapitel vorgestellten Daten. Diese Daten können zudem zur Weiterverarbeitung in PeTrack auch eingelesen werden und bieten damit die Möglichkeit, sitzungsübergreifend Laufwege zu bearbeiten. Rechts stehen die für die anschließende Bewegungsanalyse (siehe Abschnitt 4.3) aus diesen Daten durch die Positionsbestimmung ermittelten relevanten Informationen wie Personennummer, Framenummer, 3D-Position und Kopfausrichtung in der Bewegungsebene, so wie sie in diesem Abschnitt bestimmt werden sollen.

Bereits in Abschnitt 3.1.1 wurde die Vermessung im Bild eingeführt. Aus Gleichung 3.2 und 3.4 ergibt sich für die orthogonale Sicht aus der Pixelposition $(u \ v)^T$ die Position im Raum von

$$\vec{x} = R^{-1}(\vec{x}' - \vec{t}) = R^{-1}\left(\frac{z_0}{f} \begin{pmatrix} \frac{1}{s_x}(u - c_x) \\ \frac{1}{s_y}(v - c_y) \\ 1 \end{pmatrix} - \vec{t}\right). \quad (3.18)$$

Dabei ist R und \vec{t} durch die Orientierung und Position der Kamera im Raum gegeben und f, s_x, s_y, c_x und c_y ergeben sich aus den Kameraparametern. Zur Bestimmung der realen Position $p_i(t)$ aus der

...	
0 -100000 0 1999 920 86 60 64 2000	
1238.04 648.42 -1 -1 -1 100 1231.37 649.43 226 203 209	1 0 243.31 -6.79 173 -0.6217 0.7832
1236.52 648.55 -1 -1 -1 100 1229.77 649.50 219 201 204	1 1 243.43 -6.77 173 -0.6369 0.7708
1235.05 648.56 -1 -1 -1 100 1228.51 649.33 220 194 200	1 2 243.53 -6.60 173 -0.6593 0.7518
1233.62 648.88 -1 -1 -1 100 1227.21 649.52 215 197 203	1 3 243.46 -6.49 173 -0.6290 0.7773
1232.34 649.01 -1 -1 -1 70 0 0 -1 -1 -1	1 4 243.47 -6.41 173 -0.6355 0.7720
1230.97 649.45 -1 -1 -1 68 0 0 -1 -1 -1	1 5 243.41 -6.25 173 -0.6478 0.7617
1229.72 650.24 -1 -1 -1 100 1223.16 649.67 223 202 209	1 6 243.49 -6.14 173 -0.6598 0.7514
1228.61 650.99 -1 -1 -1 66 0 0 -1 -1 -1	1 7 243.51 -6.28 173 -0.6510 0.7590
1227.09 652.09 -1 -1 -1 100 1220.51 650.78 229 190 199	1 8 243.56 -6.17 173 -0.6573 0.7535
1225.39 652.88 -1 -1 -1 100 1218.83 651.42 229 204 208	1 9 243.56 -6.06 173 -0.6571 0.7538
1224.15 653.65 -1 -1 -1 100 1217.84 651.65 222 179 188	1 10 243.51 -6.04 173 -0.6581 0.7528
1222.57 654.50 -1 -1 -1 100 1216.51 652.09 214 172 173	1 11 243.50 -5.96 173 -0.6601 0.7511
1221.04 655.70 -1 -1 -1 100 1215.05 652.85 235 171 171	1 12 243.30 -5.84 173 -0.6578 0.7531
1219.46 657.11 -1 -1 -1 100 1213.92 653.72 232 187 188	1 13 243.26 -5.90 173 -0.6497 0.7601
...	...
0 -100000 0 1554 1510 88 48 51 1555	
1609.26 549.80 -1 -1 -1 100 1603.25 549.91 169 130 129	2 7 126.86 -2.94 178 -0.9993 -0.0360
1609.71 549.75 -1 -1 -1 100 1603.12 549.81 165 109 113	2 8 126.84 -2.73 178 -0.9986 -0.0522
1608.72 549.40 -1 -1 -1 100 1602.58 549.58 168 111 114	2 9 126.78 -2.56 178 -0.9986 -0.0528
1608.24 549.20 -1 -1 -1 100 1602.11 549.39 188 146 147	2 10 126.82 -2.46 178 -0.9993 -0.0356
1607.78 549.18 -1 -1 -1 100 1601.85 549.56 160 108 111	2 11 126.84 -2.22 178 -0.9998 -0.0193
1607.06 549.04 -1 -1 -1 100 1600.98 549.44 184 137 140	2 12 126.92 -2.37 178 -0.9991 -0.0412
...	...

Abb. 3.30: In PeTrack exportierte Datenstrukturen der Laufwege: die Zeilen beinhalten (links) 2D-Pixelposition der Person, hier nicht vorliegende Disparitäts-Informationen, Trackingqualität, 2D-Pixelposition und Farbe des Farbmarkerelementes; (rechts) Personennummer, Framenummer, 3D-Position und Kopfausrichtung der Person

Pixelposition $p'_i(t) = (u \ v)^T$ ist somit nur noch der Abstand z'_0 des Objektes von der Kamera nötig (siehe Abb. 3.5). Die Personengröße $p_i^h \in \mathbb{R}$ ergibt sich in ebenen Bewegungen aus der Aufhängehöhe der Kamera a und dem Abstand z'_0 :

$$p_i^h = a - z'_0. \quad (3.19)$$

Sollte die individuelle Personengröße p_i^h für die Positionsbestimmung außer Acht gelassen werden, ergibt sich aufgrund der perspektivischen Verzerrung bei einer Größendifferenz d_h zweier Personen ein Fehler e_h in Abhängigkeit vom Winkel α zum Kameralot bzw. der optischen Achse von (siehe Abb. 3.31 a):

$$e_h = |d_h \tan \alpha|. \quad (3.20)$$

Würde z. B. bei der Erfassung von Personen mit Größen zwischen 160 cm und 200 cm immer eine Durchschnittsgröße von $\bar{p}^h = 180$ cm angenommen werden, ergäbe sich in der Mitte des horizontalen Randes eines Weitwinkelobjektives bei $\alpha = 35^\circ$ ein Fehler in der Positionsbestimmung in der Ebene von

$$e_h = |(200 - 180) \text{ cm} \tan 35^\circ| \approx 14 \text{ cm}.$$

Abb. 3.31 illustriert zum einen den Positionsfehler und zeigt zudem wie dieser bei steigender Größendifferenz (z. B. Berücksichtigung von Kindern) und zum Rand des Bildes insbesondere bei Weitwinkel-aufnahmen (beliebt aufgrund des großen Erfassungsbereiches, siehe Tab. 4.2) stark ansteigt.

In den folgenden beiden Abschnitten werden zwei vom Autor verwandte Verfahren vorgestellt, die der Ermittlung der Personengröße bzw. des Abstandes zur Kamera dienen und damit den Fehler durch die perspektivische Verzerrung verringern. Weitere Verfahren wären denkbar: z. B. wurde bei den BaSiGo-Versuchen (Nr. 9) ein individueller Code zur Identifikation jeder Person genutzt, so dass einer den Code tragenden Person die vorab erfasste Größe (z. B. durch Fragebogen) zugeordnet werden kann. Eine Vereinzelung der Personen beim Eintritt in den erfassten Versuchsaufbau könnte durch

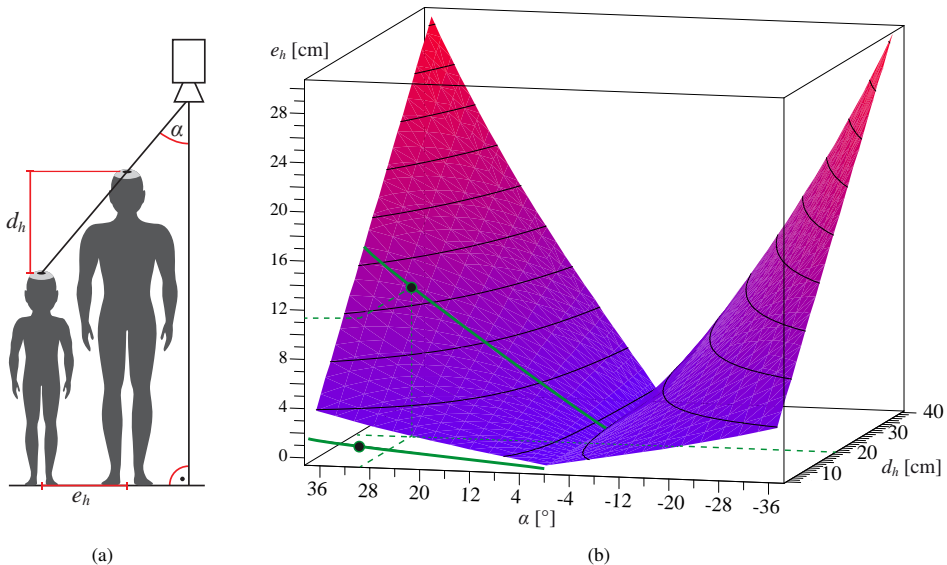


Abb. 3.31: Ortsfehler in der Bewegungsebene aufgrund unterschiedlicher Personengrößen; (a) der gleiche Punkt im senkrecht aufgenommenen Kamerabild kann aufgrund der perspektivischen Verzerrung Personen gehören, die $e_h = |d_h \tan \alpha|$ auseinander stehen, wenn ihre Größendifferenz d_h beträgt; (b) die Fläche zeigt die Abhängigkeit der ermittelten Position einer Person von der Größendifferenz der Personen und dem Winkel zum Kameralot; die steile grüne Linie zeigt den vom Kameralot weg ansteigenden Fehler für einen Kasernen-Versuch (Nr. 3) und die flachere Linie den Fehler nach Einführung der Größenunterteilung durch Farbmarkerelemente; die schwarzen Punkte zeigen den Fehler in der Mitte des äußeren horizontalen Randes des Kamerabildes

seitlich angebrachte, zeitlich synchronisierte Apparaturen wie Lichtschranken oder Kameras ebenfalls die individuelle Personengröße ermitteln. Ggf. ist auch das geschwindigkeitsabhängige Schwingverhalten (u. a. Amplitude, Frequenz) in der Personenbewegung ebenso größenabhängig und könnte zur Schätzung der Größe einer Person herangezogen werden.

3.4.1 Farbkodierung

Bei den Kasernen-Versuchen (Nr. 3 in Tab. 4.1) wurden aufgrund der Deckenhöhe (siehe Abschnitt 4.1.1.10) und des abzudeckenden Aufnahmebereiches Kameras mit einem maximalen horizontalen Öffnungswinkel von 59° benutzt (siehe Tab. 4.2, Abschnitt 4.1.1.3 und Gleichung 4.9). Die eingesetzten Probanden besaßen eine maximale Größendifferenz von 40 cm. Um den daraus resultierenden maximalen Fehler von $|40/2 \text{ cm} \tan 59^\circ/2| \approx 11,3 \text{ cm}$ zum Rand hin zu verringern, wurden die Probanden ihrer Größe entsprechend markiert. In den äußeren Ecken des entzerrten kissenförmigen Bildes (siehe Abb. 3.33) läge der maximale Fehler sogar bei 18 cm.

Der Kasernen-Marker (Nr. 1) beinhaltet ein farbiges Markerelement (siehe Abschnitt 3.2.1, Abb. 3.8 Zeile 1 und Abb. 3.13 c), das eine Unterteilung der Personengröße in 5-cm-Abschnitte erlaubte. Die Größenunterteilung verringert den maximalen Fehler damit in der Mitte des äußeren horizontalen

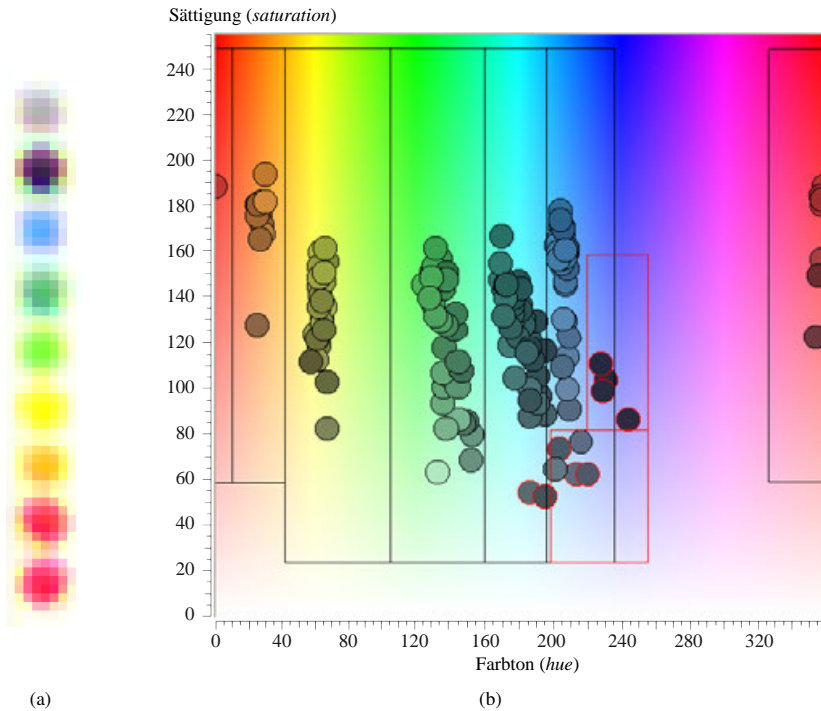


Abb. 3.32: Detektierte Farben des Farbmarkerelementes vom Kasernen-Marker (Nr. 1); (a) Aufnahme aller zehn unterschiedlichen Farbpunkte mit der in Versuchsreihe Nr. 3 verwendeten Kamera, Kameraabstand und Beleuchtung (weiß nicht sichtbar); (b) von allen Personen eines Versuches ermittelte durchschnittliche Farben als Kreisfarbe in der H-S-Ebene des HSV-Farbraumes und Rechtecke für die Zuordnung von Farben zu Größenintervallen; rot umrandeten Kreise werden mittels ihres Helligkeitswertes einem Größenintervall zugeordnet und liegen in der H-V-Ebene in den roten Rechtecken

Randes auf (siehe Abb. 3.31)

$$e_h = \left| \frac{5}{2} \text{ cm} \cdot \tan \frac{59^\circ}{2} \right| \approx 1,4 \text{ cm}.$$

In den äußeren Ecken liegt dieser maximale Fehler aufgrund nicht unterscheidbarer Personengrößen bei 2,2 cm.

Abb. 3.32 a zeigt eine Aufnahme aller zehn verwendeten Farbpunkte zur Klassifizierung der Personengröße mit der in Versuchsreihe Nr. 3 verwendeten Kamera, Kameraabstand und Beleuchtung (weiß ist auf weißem Untergrund nicht sichtbar).

In Abb. 3.33 ist die bei jedem Erkennungsschritt ermittelte Farbe entlang eines Laufweges einer Person zu erkennen, deren Farbton auch im relevanten Bildbereich stabil bleibt. An Orten mit nicht farbig gefüllten roten Kreisen wurde der Marker nicht detektiert, sondern die Person nur getrackt. Alle ermittelten Farben einer Person werden gemittelt und sind in *PeTrack* in den Farbräumen RGB und HSV ($H \hat{=}$ Farbton (*Hue*), $S \hat{=}$ Sättigung (*Saturation*), $V \hat{=}$ Helligkeit (*Value*)) in Ebenen zweier Primärvalenzen gruppierbar. Abb. 3.32 b zeigt die von allen Personen eines Versuches ermittelten durchschnittlichen Farben als Kreisfarbe in der H-S-Ebene des HSV-Farbraumes wie sie in *PeTrack*

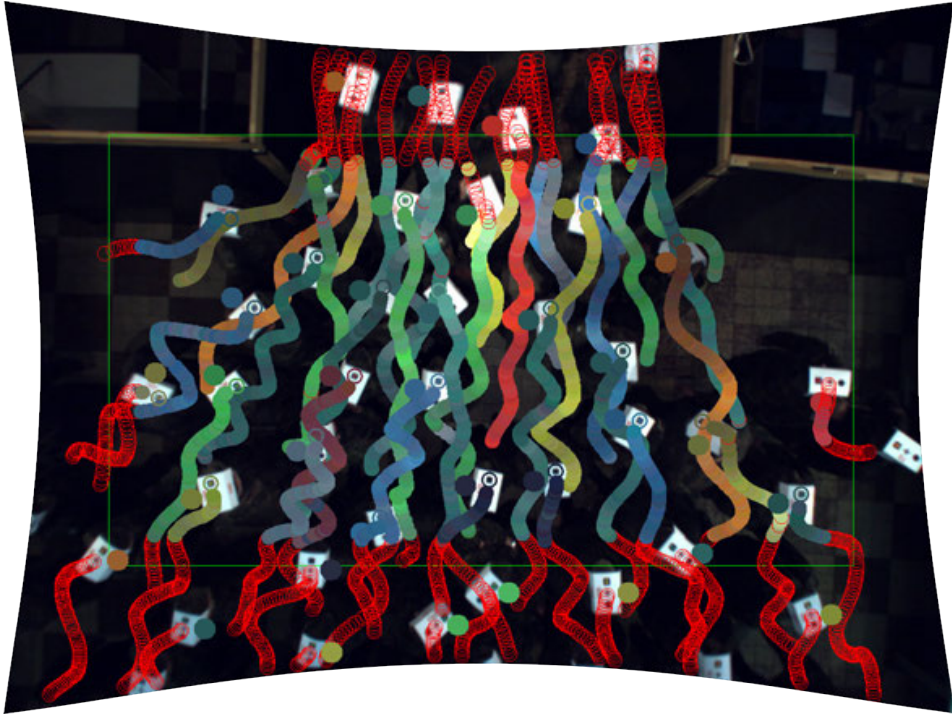


Abb. 3.33: Engstellenversuch aus Versuchsreihe Nr. 3; die Position jeder Person in den Einzelbildern der letzten vier Sekunden wird durch Kreise angezeigt; rote leere Kreise zeigen Positionen, die nur durch Tracking bestimmt wurden; die Farbbestimmung wurde innerhalb des relevanten Bereiches (grünes Rechteck) in jedem Einzelbild mit der Detektierung durchgeführt; farbige Kreise entlang des Laufweges zeigen die bestimmten Farben; im aktuellen Bild ist das Farbmarkerelement mit einem Kreis in der detektierten Farbe umrandet; die Durchschnittsfarbe wird links neben dem Kopf jeder Person als größerer Kreis angezeigt

zur Einstellung der Farbzuzuordnung zu einer Größe angeboten wird. Diese Ebene hat sich neben der H-V-Ebene als gute Wahl für die Gruppierung der erkannten Farben herausgestellt. Rechtecke dienen der Zuordnung von Farbpunkten in diesem Farbbereich zu einer bestimmten Personengröße. Die saubere Separation für echt farbige Punkte ist gut zu erkennen. Lediglich die beiden Rottöne liegen nah beieinander. Daher wurde bei der Zuordnung einer Farbe zu einer Größe darauf geachtet, dass ähnliche Farben auch ähnliche Größen repräsentieren, um im Falle einer Fehlzuordnung den Fehler möglichst gering zu halten (siehe Abb. 3.34).

Die zusätzliche Nutzung von reinen Helligkeitswerten (weißer, grauer und schwarzer Farbmarker) zur Separation hat sich als störend herausgestellt, da durch leicht unterschiedliche Kopfneigung bzw. Sitz des Markers und damit einfallendem Licht der Farbton sich im HSV-Modell verschiebt. Ein in *Pe-Track* wählbarer automatischer Weißabgleich, der die Farbe des Markers im Farbraum entsprechend so bewegt, dass bei gleicher relativer Verschiebung der Weißton auf der Pappe um das Farbmarkerelement auf einen Grauwert verschoben wird, konnte das Problem nur geringfügig lindern. Auch die Verwendung von beliebigen, unterschiedlichen Strahlern bei den Kasernen-Versuchen (Nr. 3) zur Ausleuchtung, die nicht alle die gleiche Farbtemperatur besaßen, führte zur Verschiebung von Farbtönen

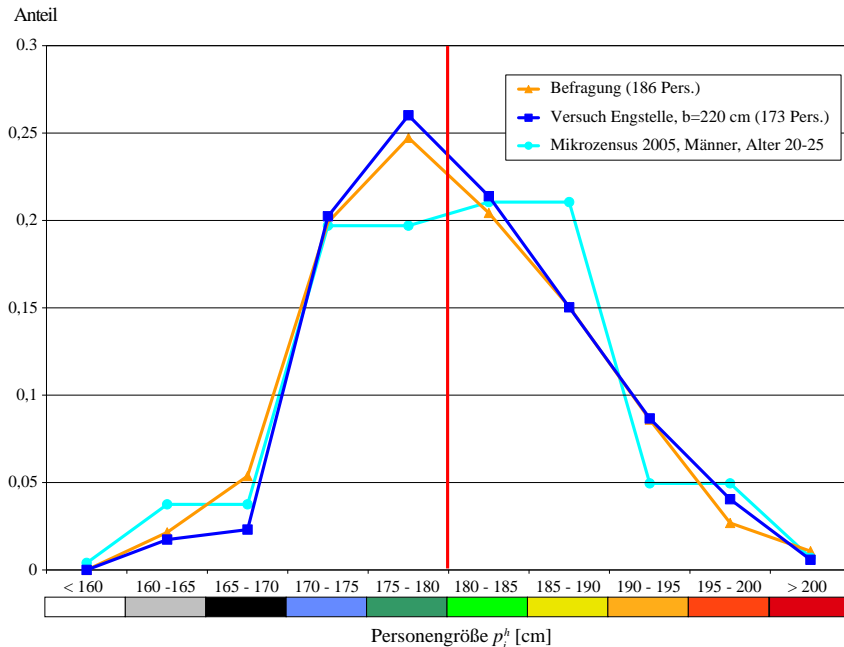


Abb. 3.34: Verteilung der Personengröße p_i^h laut Befragung aller an diesem Tag beteiligten Personen (vornehmlich männliche Soldaten), laut der ermittelten Farbzurordnung und laut des entsprechenden Mikrozensus des Jahres 2005; die das jeweilige Größenintervall repräsentierende Farbe ist unterhalb der Größenachse angegeben

insbesondere bei grauwertigen Markern. Daher werden Farbmarkerelemente mit geringem H-Wert in PeTrack getrennt behandelt und gemäß ihren V-Werten zugeordnet (rot umrandete Farbpunkte in Abb. 3.32 b).

Hingegen sorgte der Bayer-Filter aller farbigen 1-Chip-Industriekameras bei der Farbbestimmung für keine Probleme, da Farbmarker auf ein schwarzes Quadrat aufgebracht wurden (siehe Abb. 3.8 Zeile 1) und als Farbe die hellste im inneren $3\text{px} \times 3\text{px}$ Bereich ausgewählt wurde. Die einzig im Übergang vom Farbpunkt zum schwarzen Quadrat auftretende Fehlinterpolation aufgrund des Filters führt jedoch zu dunkleren Farben (im Gegensatz zu den zur Illustration auf einem weißen Untergrund aufgetragenen und aufgezeichneten Farbpunkte in Abb. 3.32 a).

Die Größenverteilung wurde mittels des farbigen Markerelementes gut getroffen. Abb. 3.34 zeigt die vergleichende Verteilung der ermittelten Größenbereiche aufgrund der Farbmarkerelemente eines Versuches und der Größen aufgrund der Befragung aller an diesem Tag teilgenommenen Probanden.

3.4.2 Disparität

Bei den Hermes-Versuchen (Nr. 6) sollten auch Laufwege von Personen in unebenen Fußgängeranlagen, wie Treppen und Tribünen, ermittelt werden. Ohne für alle Wege die exakte 3D-Geometrie und Lage zur Kamera zu kennen, wäre alleine mit dem Wissen der Personengröße keine Positionsbestimmung möglich.

Daher musste ein 3D-Messverfahren eingesetzt werden. Dieses durfte die Umgebung nicht dahingehend beeinflussen, dass die Wahrnehmung der Probanden beeinträchtigt wird, z. B. durch projizierte Muster im sichtbaren Licht. Zudem konnte kein auf Infrarotlicht basierendes Verfahren eingesetzt werden, da die Messungen auch im Freien bei Infrarotlicht aussendender Sonne stattgefunden haben. Deshalb wurden zur Entfernungs- und damit 3D-Positionsbestimmung Stereokameras eingesetzt. Weitere Verfahren und ihre Vor- und Nachteile können Abschnitt 2.3.6 entnommen werden.

Die bis dahin eingesetzten monokularen Kameras bilden eine dreidimensionale Szene in Zentralprojektion auf einen zweidimensionalen Detektorchip ab. Dabei geht die Tiefeninformation verloren. Wird die Szene jedoch mit zwei Kameras aufgenommen, können die 3D-Koordinaten derjenigen Punkte im Raum mittels Triangulation ermittelt werden, die in beiden Bildern zu erkennen und einander zuordenbar sind. Diese binokulare Sicht zur 3D-Rekonstruktion wird auch als stereoskopisches Sehen, Stereopsis oder Stereoanalyse bezeichnet. Dabei stehen für gewöhnlich die Perspektiven beider Kameras fest, so dass nicht zusätzlich die Parallaxe wie bei den Augen, sondern nur die Disparität (*disparity*), also der Unterschied zwischen den Bildern, Aufschluss über die Entfernung zum Kamerasystem gibt.

Für die Korrespondenzanalyse ist eine Identifikation von den gleichen Ort beschreibenden Punktepaa- ren in beiden Perspektiven nötig, so dass das Verfahren nur bei texturierten Oberflächen funktioniert. Dabei sind es hier, im Gegensatz zur Bestimmung von Korrespondenzen beim optischen Fluss oder auch beim Tracking (siehe Abschnitt 3.3), nicht sich zeitlich sondern örtlich leicht unterscheidende Bilder. Die Texturierung ist bei Menschen bzw. Menschenmengen für gewöhnlich gegeben. Durch zur Erkennung weiterhin eingesetzte, kontrastreiche Marker wird die Texturierung verstärkt. Natürlicher- weise funktioniert dieses Verfahren nur bei ausreichender Beleuchtung, so dass es nicht im Dunkeln ohne künstliche ggf. infrarote Zusatzbeleuchtung eingesetzt werden kann.

Wird außer für die mittlere Kopfposition wie hier die Disparität auch für andere Bereiche der Szene gebraucht, können weitere Probleme bei der Korrespondenzanalyse auftreten. Für die Subtraktion des Hintergrundes bei der markerlosen Detektion ist z. B. der Höhenverlauf der Umgebung nötig. Dabei existieren Böden oder Wände, die strukturlos sind, für die somit keine Disparität bestimmt werden kann. Die Lösung für dieses Problem ist in Abschnitt 3.5.1 beschrieben.

Für periodisch oder wiederholt auftretende Strukturen können keine eindeutigen Disparitäten be- stimmt bzw. können falsche Korrespondenzen ermittelt werden. Diese Art von Strukturen treten in der Natur und damit in Menschenmengen selten auf. Bei anthropogenen Elementen wie den zu durchlau- fenden künstlichen Bauten kommen solche Wiederholungen eher vor. Eine Lösung für solche Stellen ist ebenfalls Abschnitt 3.5.1 zu entnehmen. Dass die Marker nicht als Wiederholung gesehen werden, liegt an der weiter unten beschriebenen Epipolarbedingung, die gleiche Pixel nur entlang einer Zeile suchen lässt. Zudem sollten, wie ebenfalls weiter unten beschrieben, die Grenzen des Suchbereichs eingeschränkt werden.

Alle Aufnahmen werden überkopf durchgeführt, so dass alle Markerelemente der Personen gut er- kennbar sind (siehe Abschnitt 4.1.1.1). Dadurch ist ebenfalls gewährleistet, dass beide Perspektiven eines Stereokamerasystems mit einem kleinen Kameraabstand diese Elemente zur Ermittlung der Dis- parität auf dem Kopf sehen. Der Abstand der optischen Zentren der Kameras wird Basislänge (*base- line*) genannt und betrug bei den Hermes-Versuchen (Nr. 6 in Tab. 4.1) mit der Stereokamera Nr. 4 (siehe Tab. 4.2) $b = 24\text{ cm}$. Nicht alle Punkte einer Szene sind zum einen aufgrund von Verdeckung oder zum anderen aufgrund der Kameraausrichtung von beiden Perspektiven aus einsehbar. Siehe z. B. die Überlappungsbereiche der Kameras für den abgebildeten Kopf in Abb. 3.35 oder den nicht in Abb. 3.36 a bzw. c enthaltenen rechten Bildrand von Abb. 3.36 b bzw. d und damit den breiteren

Streifen am rechten Rand fehlender farbig überlagter Disparität in Abb. 3.39.

Dies spielt für die Größenermittlung keine Rolle, jedoch für das später eingeführte Verfahren zur markerlosen Erkennung. Wie damit umgegangen wird, kann Abschnitt 3.5.1 entnommen werden.

Für einen guten Kompromiss zwischen geringer Verdeckung beider Perspektiven und einem großen Intervall von Pixelabständen für den interessierten Tiefenbereich ist der Abstand der Kameras im Stereokamerasystem entscheidend. Ein großes Intervall von Pixelabständen bzw. ein großer Disparitätsbereich entspricht mehr Quantisierungsstufen bzw. einem geringeren Quantisierungsfehler für die Tiefe und somit einer höheren Genauigkeit in der Bestimmung der Personengröße. In Abb. 3.35 ist dieser Unterschied illustriert, wie sich durch eine Vergrößerung der Basislänge ($b_2 > b_1$) ein größerer Disparitätsbereich ergibt. Die absolute Disparität

$$\delta_{b_1}(u_1) = u_2 - u_1 < u_3 - u_1 = \delta_{b_2}(u_1)$$

ist für b_2 größer und ebenso der Disparitätsbereich

$$|\delta_{b_1}(u_1) - \delta_{b_1}(u'_1)| < |\delta_{b_2}(u_1) - \delta_{b_2}(u'_1)|.$$

Die Berechnung der Disparität und die Tiefenabdeckung eines Pixelabstandes wird weiter unten ausgeführt.

Zur Bestimmung der Disparität sind einige Vorverarbeitungsschritte nötig oder vereinfachen die Bestimmung des Pixelabstandes:

1. Entzerrung beider Kameraperspektiven entsprechend Abschnitt 3.1.2,
2. Ausrichtung der Ansichten, so dass die optische Achse beider Perspektiven parallel zueinander liegen,
3. Ausrichtung der Ansichten, so dass dieselben Bildzeilen beider Perspektiven die gleiche reale Ebene beschreiben.

Die Epipolargeometrie (*epipolar geometry*), die die geometrische Beziehung zwischen den beiden Ansichten des Stereokamerasystems beschreibt, erlaubt die Ausrichtung gemäß den Punkten 2 und 3 durch Rektifizierung (*rectification*) nach Bestimmung der Fundamentalmatrix (*fundamental matrix*) [457] (siehe Abb. 3.37). Ein Ergebnis ist beispielhaft in Abb. 3.36 zu sehen. Die beiden direkt vom Sensor gelieferten Ansichten a und b werden durch die Rektifizierung in c und d überführt. Die nachträglich eingefügten horizontalen roten Linien sollen zeigen, dass die vorher nicht ausgerichteten Ansichten aus a und b in c und d so transformiert wurden, dass gleiche Bildzeilen die gleiche,

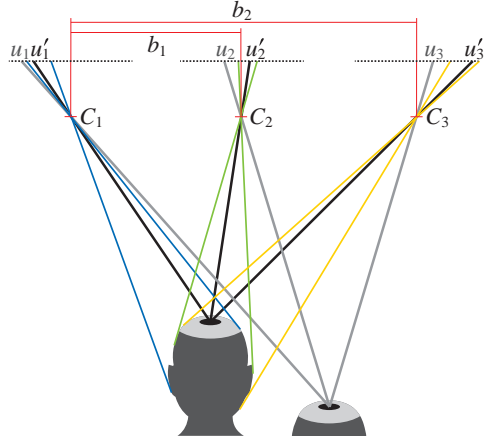


Abb. 3.35: Einfluss der Basislänge b_* der Perspektiven eines achsparallelen Stereokamerasystems auf die Größe des gemeinsam sichtbaren Bereiches (farbige dünne Geradenpaare zeigen Sichtbereich des Kopfes) und auf die Größe der Disparität und des Disparitätsbereiches (schwarze und graue Geraden zeigen tiefenabhängige Disparität); alle Geraden bzw. Sehstrahlen gehen durch die optischen Zentren C_* und fallen auf den Sensor an Pixelkoordinate u_*

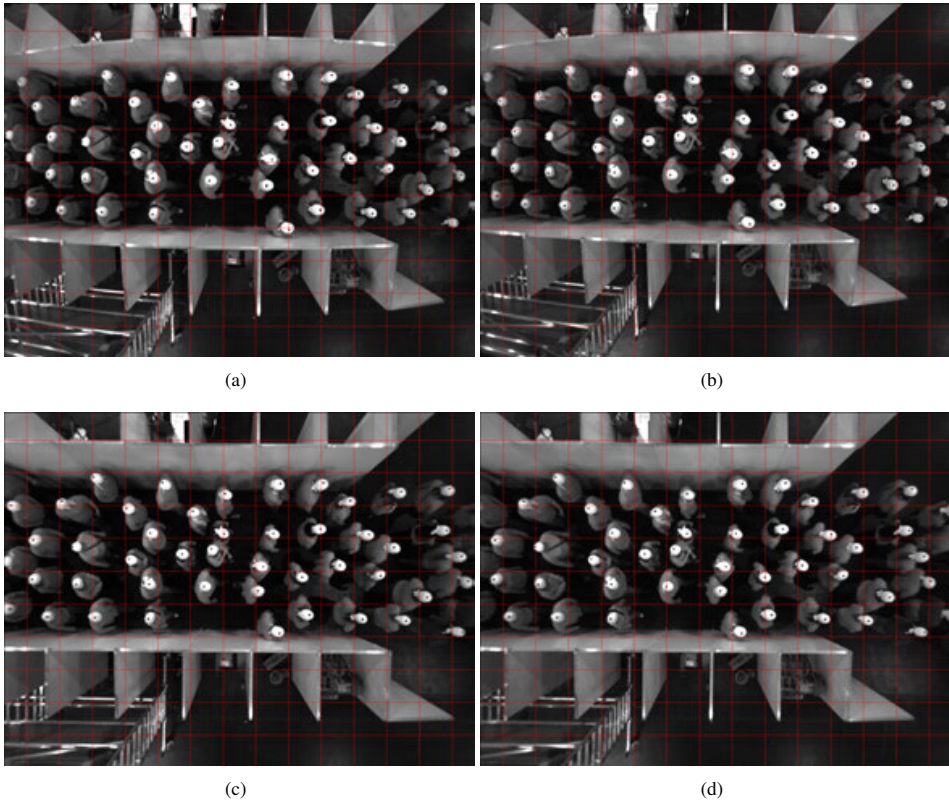


Abb. 3.36: Experiment aus Versuchsreihe Nr. 6; (a, b) beide versetzten Blickwinkel der eingesetzten Stereokamera Nr. 4 (siehe Tab. 4.2); (c) das entzerrte und rektifizierte linke Bild a; (d) das entzerrte und rektifizierte rechte Bild b (siehe Abschnitt 3.1.2); die roten Linien sollen das Ergebnis der Rektifizierung verdeutlichen und insbesondere zeigen, dass gleiche Punkte der Szene auf dieselben Bildzeilen abgebildet werden

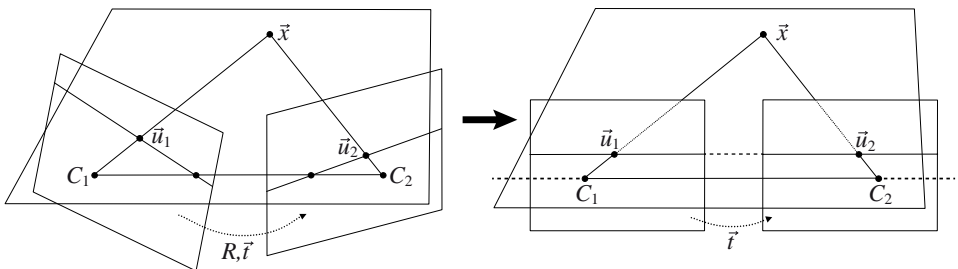


Abb. 3.37: Rektifizierung, so dass optische Achsen parallel liegen und reale Punkte auf dieselben Bildzeilen projiziert werden

durch die zwei optischen Zentren gehende Ebene in der Realität beschreiben. Die vertikalen roten Linien erlauben eine bessere Abschätzung der Disparitäten zwischen den beiden Ansichten c und d. Die Transformation der Ansichten, so dass Bildzeilen gleiche Ebenen im Raum beschreiben, dient der einfacheren anschließenden Korrespondenzanalyse, da nun gleiche Punkte nur noch entlang von Bildzeilen und nicht mehr im gesamten Bild gesucht werden müssen.

Nach der Rektifizierung gelten nun folgende Eigenschaften für die Korrespondenzanalyse:

- Epipolarbedingung (*epipolar constraint*): ein Punkt im Raum liegt in beiden Ansichten auf einer Bildzeile; dieser Punkt liegt in der rechten Ansicht weiter links innerhalb der Bildzeile als in der linken Ansicht,
- Eindeutigkeitsbedingung (*uniqueness constraint*): ein Punkt im Raum kann maximal nur zu einer Abbildung in beiden Ansichten führen,
- Glattheitsbedingung (*smoothness constraint*): größtenteils ändert sich die Disparität kontinuierlich, nur bei Objektübergängen die hintereinander liegen gibt es Tiefensprünge,
- Reihenfolge-Bedingung (*ordering constraint*): bei kleiner Basislänge bleibt für gewöhnlich die Reihenfolge der Punkte zwischen den Perspektiven entlang der Bildzeilen erhalten.

Bei der Korrespondenzanalyse können zwei Klassen von Verfahren zur Findung der Korrespondenzen unterschieden werden: Verfahren, die ähnliche Pixel oder Pixelblöcke suchen (Klasse I) oder Verfahren, die in einem Vorverarbeitungsschritt Bildmerkmale wie Eckpunkte, Kanten, Regionen oder Objekte in beiden Ansichten bestimmen und den Abstand dieser Merkmale heranziehen (Klasse II). Oft werden auch Elemente aus beiden Klassen kombiniert. In `PeTrack` wurden vier Verfahren integriert:

ptGrey : ein vom Kamerahersteller der Stereokameras Nr. 3 und 4 über die Bibliothek Triclops [474] bereitgestelltes Verfahren; es basiert auf einem im Übersichtsartikel [475] empfohlenen Korrespondenzverfahren ähnlich dem nachfolgend aufgeführten BM-Verfahren und kann zudem zur Plausibilitätsprüfung Algorithmen zur Güte der Textur, zur Eindeutigkeit, zur Vermeidung von Zacken im Disparitätsbild [476] und zur Rückwärtssuche durch Tausch der Ursprungsbilder [477] nutzen; in `PeTrack` werden nur die beiden letztgenannten Prüfungen durchgeführt; eine in diesem Verfahren wählbare Kantentreue kann in `PeTrack` hinzugeschaltet und deren Maskengröße angepasst werden (Klasse I, durch Aktivierung des Kantentests auch Klasse II),

OpenCV block matching : (kurz BM-Verfahren) in Bibliothek OpenCV [446,447] implementiertes, einfaches und schnelles Verfahren basierend auf [478,479], bei dem im Kern bei der Korrespondenzsuche das Minimum der Summe der absoluten Differenzen in den Bildern beider Perspektiven B_1 und B_2 in einem Vergleichsfenster der Größe $2\omega + 1$ um \vec{u}_1 bzw. um $\vec{u}_1 + (\delta, 0)$ gesucht wird (Klasse I):

$$\operatorname{argmin}_{\delta} \sum_{\omega_x=-\omega}^{\omega} \sum_{\omega_y=-\omega}^{\omega} (B_1(\vec{u}_1 + (\omega_x \ \omega_y)^T) - B_2(\vec{u}_1 + (\delta \ 0)^T + (\omega_x \ \omega_y)^T))^2, \quad (3.21)$$

OpenCV semi-global block matching : (kurz SGBM-Verfahren) in Bibliothek OpenCV implementiertes Verfahren basierend auf [480], bei dem für die Korrespondenzsuche eine globale Kostenfunktion gemäß [481] minimiert wird; dabei wird diese in horizontaler, vertikaler und diagonalen Richtung akkumuliert und mit einem kantenerhaltenden Glattheitskriterium verknüpft [482] (Klasse I und II),

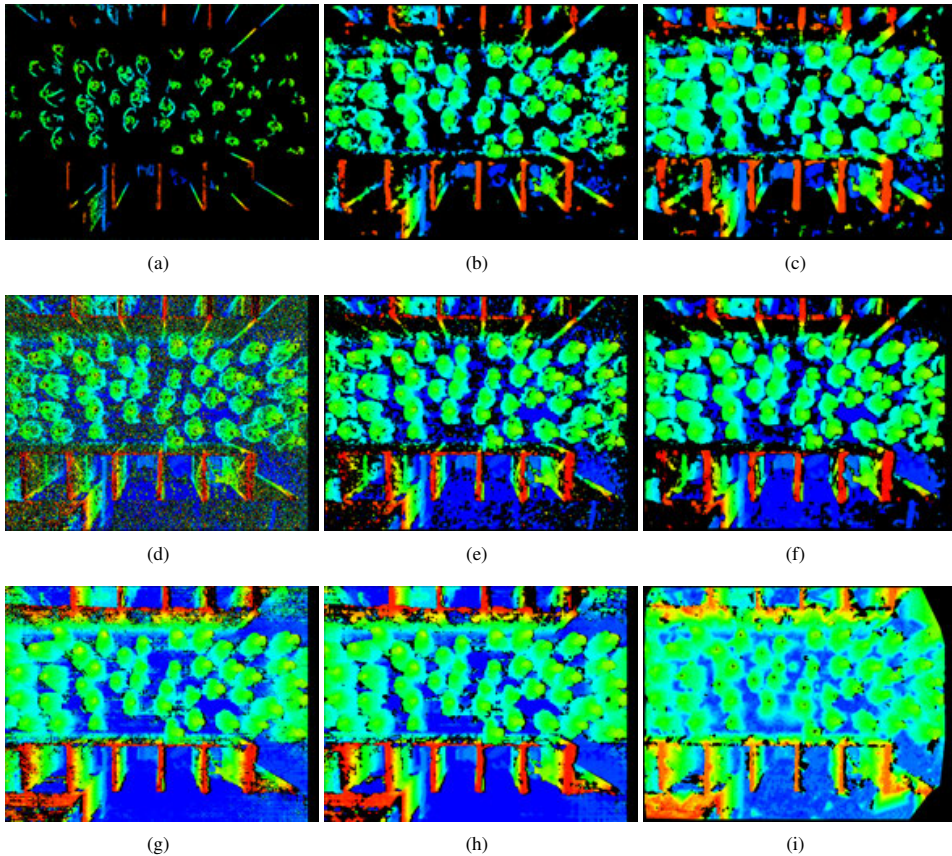


Abb. 3.38: Farbkodierte Disparitäten des Bildes aus Abb. 3.36 bestimmt durch die in *PeTrack* implementierten Verfahren gemäß Bezeichnung aus der Liste auf Seite 91: (a-c) *ptGrey*-Verfahren mit Kantenmaske von 5 px und Vergleichsfenstergröße von (a) 3 px, (b) 13 px, (c) 23 px; (d-f) *BM*-Verfahren mit Fenstergröße von (d) 5 px, (e) 13 px, (f) 21 px; (g, h) *SGBM*-Verfahren mit Fenstergröße von (g) 3 px, (h) 11 px; (i) *ELAS*-Verfahren; die Farbskala entspricht bei allen Bildern der aus Abb. 3.40

efficient large-scale stereo matching : (kurz *ELAS*-Verfahren) in Bibliothek *libelas* (*LIB*rary for *E*fficient *L*arge-scale *S*tereo matching) [483] implementiertes Verfahren, das robuste Bildmerkmale sucht, um die Disparität im Vorverarbeitungsschritt zu schätzen und zugleich den Wertebereich der Disparität automatisch zu bestimmen; weitere Details in [484] (vornehmlich Klasse II).

Abb. 3.38 zeigt Disparitätsbilder dieser Verfahren mit unterschiedlichen Parametern. Dabei ist der Wertebereich der zu bestimmenden Disparität für alle Verfahren auf das Intervall zwischen 29 px und 46 px beschränkt worden, da nur diese Werte für die ebenen Versuche aus Versuchsreihe Nr. 6 (siehe Abb. 3.40) vorkommen dürfen. Nur das *ELAS*-Verfahren bestimmt diesen Bereich automatisch. Die Einschränkung des Wertebereiches beschleunigt die Bestimmung und verringert fehlerhafte Korrespondenzen, die u. a. durch wiederholende Strukturen auftreten. Eine fehlerhaft detektierte wiederho-

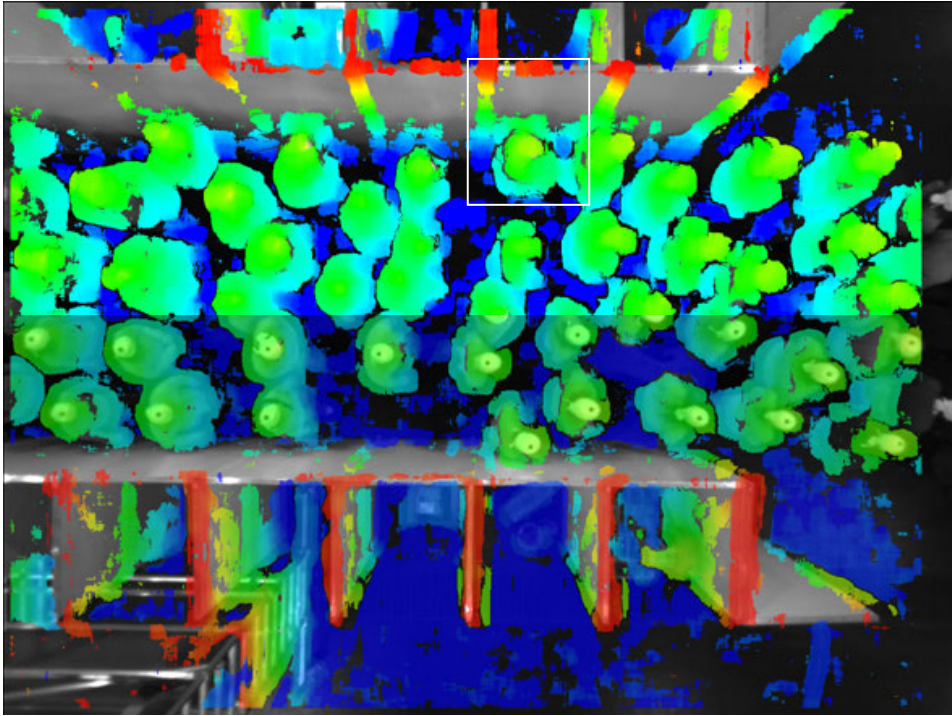


Abb. 3.39: Abb. 3.36 d mit überlagerter farbkodierter Disparität mit Verfahren und Parametern gemäß Abb. 3.38 f; die Entfernung zur Kamera beträgt 534 cm (rot) bis 784 cm (blau); im unteren Bereich ist die Disparität halbdurchsichtig dargestellt, um die Zuordnung zum Bildpixel herstellen zu können; der weiß umrandete Bereich wird in Abb. 3.40 näher betrachtet

lende Struktur in einer Bildzeile ist in Abb. 3.39 unten links zu sehen, bei der die oberen Gerüststufen nah am Boden gesehen werden.

Zwei Marker könnten aufgrund der gleichen Textur in maximaler Kopfhöhe von 200 cm und damit einem Kameraabstand von 584 cm und maximal erlaubter Disparität von 46 px fälschlich als korrespondierend angesehen werden, wenn diese laut Strahlensatz (siehe Abb. 3.5) höchstens

$$584 \text{ cm} \frac{46 \text{ px}}{983 \text{ px}} \approx 27 \text{ cm}$$

voneinander entfernt liegen. Diese Markerverwechslung bei der Korrespondenzanalyse trat bei allen Verfahren jedoch nicht auf, da es zum einen selten ist, dass zwei Personen sehr nah und dazu genau an einer Bildzeile ausgerichtet sind, und zum anderen die Reihenfolge-Bedingung (siehe Seite 91) für gewöhnlich verletzt wird.

Alle Verfahren besitzen eine Vielzahl von wählbaren Parametern. Für diese sind optimale Werte mittels der Versuchsreihe Nr. 6 heuristisch bestimmt worden. In *PeTrack* frei wählbar ist neben dem Wertebereich der Disparität lediglich die Größe des Vergleichsfensters und für das *ptGrey*-Verfahren die optionale Hinzunahme der Untersuchung der Kantentreue mit wählbarer Maskengröße. Abb. 3.38 zeigt die Auswirkung unterschiedlicher Größen des Vergleichsfensters, die für die drei erstgenannten variabel eingestellt werden können.

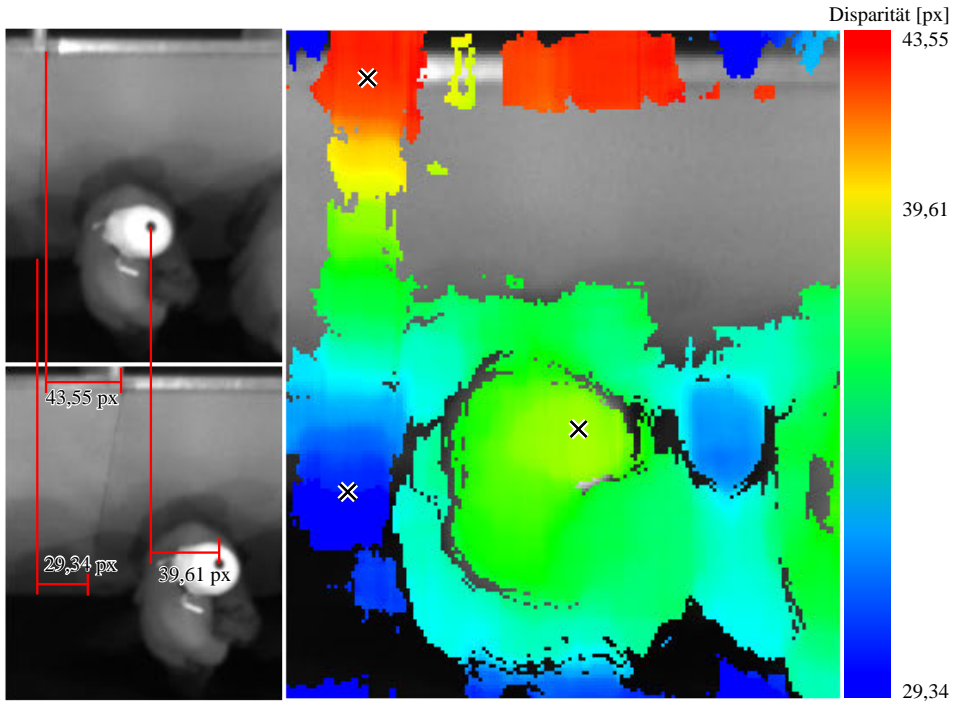


Abb. 3.40: Geometrische Bestimmung der Disparität für drei Punkte aus dem umrandeten Bereich aus Abb. 3.39; metrische Werte der Disparität in Pixeln sind Tab. 3.2 zu entnehmen

Disparität δ [px]	Entfernung zur Kamera z'_0 [cm]	Höhe über Boden $a - z'_0$ [cm]
43,55	534	250
39,61	588	196
29,34	784	0

Tab. 3.2: Metrische Größen der in Abb. 3.40 markierten Punkte, bestimmt über Gleichung 3.23 und mit den Kameraparametern $f' = 983$ px und $b = 24$ cm

Große Masken erzeugen typischerweise weniger Fehler bei der Bestimmung der Korrespondenzen und glätten das Ergebnis, lassen damit aber den Höhenverlauf unschärfer werden (siehe Abb. 3.38 d-f und Abschnitt 3.6.5).

Für die Bestimmung der Personengrößen kann jedes Verfahren gewählt werden, da der kontrastreiche Marker für alle Verfahren gleiche Disparitäten erzielte. Die Maskengröße wurde mindestens auf 11 px gesetzt, um ein leichtes Rauschen bei der Disparität zu entfernen. Für die anschließend vorgestellte markerlose Personendetektion mittels des Disparitätsbildes erzielten die Verfahren unterschiedlich gute Ergebnisse. Diese werden in Abschnitt 3.6.5 vorgestellt.

Die Laufzeit der Verfahren unterscheidet sich für eine Maskengröße von 11 px um den Faktor von 6,

wobei das ptGrey-Verfahren das schnellste und das ELAS-Verfahren das langsamste ist (für absolute Zeiten siehe Abschnitt 3.6.8). Dies ist aber für die nachgelagerte Auswertung irrelevant, sondern würde erst dann entscheidend sein, wenn eine Echtzeiterfassung durchgeführt werden soll.

In Abb. 3.38 liegen neben Pixeln in der Bildmitte insbesondere am Rand unterschiedlich breite Streifen ohne Disparitätswerte. Der breite rechte Streifen rührt daher, dass dieser Bereich nur von der rechten Perspektive aus gesehen wird und somit keine Korrespondenzen ermittelt werden können. An den restlichen Rändern steigt für das erste, zweite und vierte Verfahren der fehlende Disparitätsbereich aufgrund der wachsenden Fenstergröße, in der gemäß Gleichung 3.21 um ein Pixel herum die Farb- oder Helligkeitsdifferenzen ermittelt und somit ein Abstand von ω px zum Rand eingehalten werden muss.

Sind durch die Korrespondenzanalyse Pixel $\vec{u}_1 = (u_1 \ v_1)^T$ und $\vec{u}_2 = (u_2 \ v_2)^T$ auf Subpixel-Ebene im rechten und linken Bild des Stereokamerasystems gefunden, die den gleichen Punkt im Raum zeigen, so ergibt sich daraus die Disparität für das rechte Bild von

$$\delta(u_2) = u_1 - u_2. \quad (3.22)$$

In Pixelzeile $v_1 = v_2$ unterscheidet sich somit die Pixelposition des gleichen Punktes an Spalte u_2 im rechten Bild um $u_1 - u_2$ Pixel vom linken Bild. Man beachte, dass Abb. 3.35 und Abb. 3.41 das Modell der Lochkamera zur Veranschaulichung der Strahlenverläufe benutzen und somit die Koordinaten um die Bildzeilenmitte gespiegelt werden müssen, damit sie z. B. mit den realen Aufnahmen in Abb. 3.36 übereinstimmen. Für die Disparität bedeutet dies, dass sich das Vorzeichen umkehrt.

Abb. 3.39 zeigt die Disparität zwischen den Bildern Abb. 3.36 c und d farbig auf das rechte Bild überlagert. Für Pixel, an denen keine Farbe kodiert ist, konnte keine Disparität ermittelt werden. Dies ist insbesondere in strukturlosen Bereichen und an Objekten, die sich entlang der Bildzeile nicht verändern, der Fall. In PeTrack besteht neben der Visualisierung der Disparität über die Farbe entlang der maximal gesättigten und hellsten Farbtöne im HSV-Farbmodell auch die Möglichkeit, diese über eine Grauwertskala zu visualisieren (siehe Abb. 3.50). Zudem kann die Transparenz der überlagerten Farbe eingestellt werden, so dass die Position der Disparität im Bild besser gesehen werden kann (siehe Abb. 3.39).

In Abb. 3.40 wird für einen Ausschnitt aus Abb. 3.36 c bzw. d und Abb. 3.39 die Pixelposition und Bestimmung der Disparität beispielhaft an drei Punkten gezeigt: die untere Kante der Wand als tiefstes, den Marker des Kopfes als mittleres und die Oberkante der Wand als höchstes Element. Je weiter der Punkt von der Kamera entfernt ist, desto kleiner ist die Disparität.

Da die Korrespondenzanalyse auf eine Pixelzeile reduziert wurde, kann die Herleitung direkt aus Abb. 3.41 mittels Strahlensatz erfolgen [475, 485, 486]:

$$\frac{z'_0}{z'_0 + f'} = \frac{b}{b + \delta(u_2)} \Leftrightarrow z'_0 = \frac{bf'}{\delta(u_2)}. \quad (3.23)$$

Das Stereokamerasystem ist hier achsparallel und die Brennweite f' beider Kameras im Stereokamerasystem identisch. Die Entfernung zur

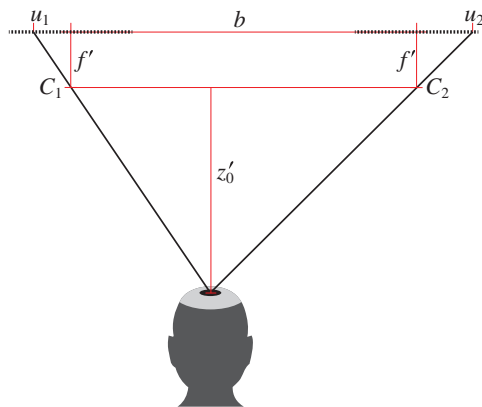


Abb. 3.41: Entfernung z'_0 eines Punktes von einem achsparallelen Stereokamerasystem ist nur abhängig von den Brennweiten f , der Basislänge b und der Disparität $\delta(u_2) = u_1 - u_2$

Kamera z'_0 parallel zur optischen Achse ist damit umgekehrt proportional zur Disparität ($\delta(u_2) \propto 1/z'_0$) und direkt proportional zur Basislänge b und Brennweite f' , so dass die absolute Sensibilität der Entfernungsbestimmung mit dem Quadrat des Abstandes abnimmt [487]. Die Disparität und Brennweite $f' = fs$ sind in Pixel angegeben, so dass der Umrechnungsfaktor s aus Abschnitt 3.1 zur Überführung in ein metrisches Maß herangezogen werden muss:

$$z'_0 = \frac{bf}{\frac{\delta(u_2)}{s}} = \frac{sbf}{\delta(u_2)}. \quad (3.24)$$

Mit der Disparität lässt sich somit direkt aus Gleichung 3.18 die Position im Raum bestimmen bzw. für die ebenen Versuche aus Gleichung 3.19 die Personengröße angeben.

Die Tiefenauflösung ist dabei, wie zu Beginn bereits erwähnt, aufgrund des umgekehrt proportionalen Zusammenhangs der Disparität und der Entfernung gröber, umso weiter man sich vom Stereokamerasystem weg bewegt, so dass die Genauigkeit der Entfernungsbestimmung abnehmen muss. Abb. 3.42 a veranschaulicht die Zunahme des ein Pixel überstreichenden Raumes und damit die Erhöhung der Entfernungsschritte bei abnehmender Disparität. Jede Raute steht dabei für ein Pixel Disparität. Die Abbildung zeigt zudem, dass eine Disparität - egal wo im Bild gemessen - immer auf dieselbe Entfernung fällt, da die zur Basislinie parallel laufenden roten Linien durch die Kreuzungspunkte gleicher Disparität verlaufen. Abb. 3.42 c zeigt die Abhängigkeit der Entfernung von der Disparität und der Basislänge und unterstreicht damit die am Anfang des Abschnittes erwähnte Verbesserung der Quantisierung der Tiefe bei steigender Basislänge. Die rote Linie bei $b = 24\text{ cm}$ und Abb. 3.42 b beschreiben den Verlauf von Gleichung 3.23 für die in den Hermes-Versuchen (Nr. 6) eingesetzte Kamera Nr. 4 mit Brennweite $f' = 983\text{ px}$. Die bei den ebenen Versuchen vorkommenden Werte aus Abb. 3.40 sind in Abb. 3.42 b hervorgehoben. Die Kamera Nr. 4 besitzt zwei mögliche Basislängen von 12 cm und 24 cm. Abb. 3.42 d zeigt die zuvor beschriebene Intervallausweitung der Disparität bei steigender Basislänge und damit den Grund für die Nutzung der Basislänge von 24 cm für die Erfassung unserer Versuche. Die kleinere Basislänge könnte für näher an der Stereokamera liegende Objekte eingesetzt werden, da diese ggf. nur so von beiden Perspektiven sichtbar sind. Zudem verringert sich generell durch eine kleine Basislänge der von nur einer Perspektive sichtbare Bereich.

3.4.3 Fehler der Disparität

Die zuvor gezeigten Disparitätsbilder zeigen je nach Verfahren unterschiedlich ausgeprägte Fehler oder Bereiche, an denen keine Disparität bestimmt werden konnte. Diese können von zum Teil bereits erwähnten Problemen herrühren [488, 489]:

1. schwach bis gar nicht texturierte Oberfläche (ggf. auch nur in Richtung der Bildzeilen),
2. zu geringe oder zu starke (Überstrahlung) Beleuchtung,
3. identische Bereiche entlang einer Bildzeile, z. B. durch periodische Anordnung von Objekten,
4. Bereiche, die nicht von beiden Kameraperspektiven gesehen werden können,
5. Eigenschaften für die Korrespondenzanalyse nach der Rektifizierung entsprechend Seite 91 (Epipolarbedingung, Eindeutigkeitsbedingung, Glattheitsbedingung, Reihenfolge-Bedingung) sind nicht erfüllt,

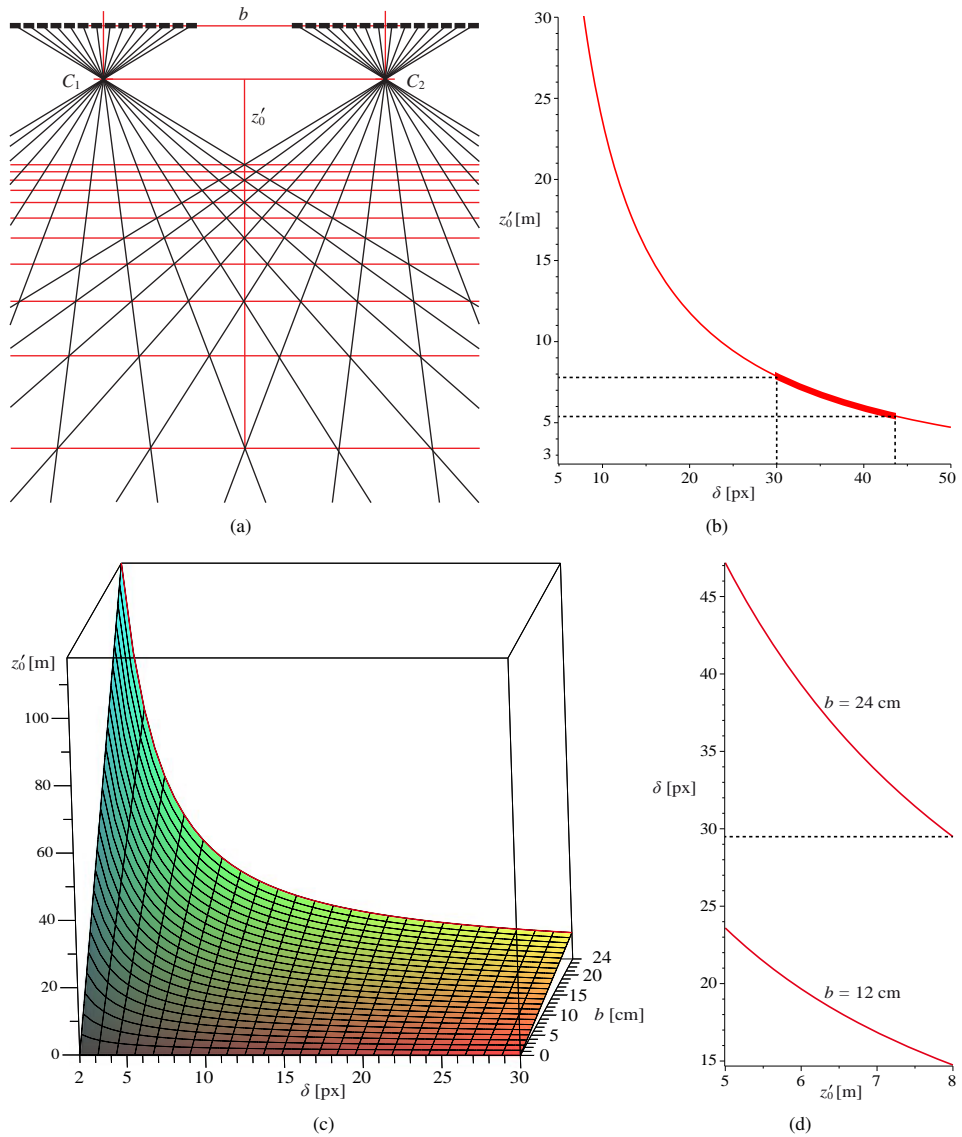


Abb. 3.42: Abhängigkeit zwischen der Entfernung zum Stereokamerasystem und der Disparität; (a) schwarze Linien sind Sehstrahlen von jedem Pixel durch die optischen Zentren C_1, C_2 des Stereokamerasystems; rote Linien parallel zur Basislinie sind die Entfernungsschritte je Pixel Disparität; (b) Graph der Gleichung 3.23 für Stereokamera Nr. 4 mit Brennweite $f' = 983$ px und Basislänge $b = 24$ cm; die für die ebenen Versuche aus Versuchsreihe Nr. 6 relevanten Disparitäten sind hervorgehoben; (c) zeigt zudem die Abhängigkeit von Gleichung 3.23 zur Basislänge; (d) Disparitäten für beide möglichen Basislängen von Kamera Nr. 4 für relevante Kameraentfernungen von Versuchsreihe Nr. 6

6. Objektoberflächen, deren Aussehen vom Blickwinkel abhängen, z. B. reflektierende Oberflächen, deren Lichtreflexionen mit dem Betrachterstandpunkt wandern, oder Oberflächen mit einer winkelabhängigen bidirektionalen Reflektanzverteilungsfunktion, so dass sich das Aussehen von identischen realen Punkten in der Szene zwischen den Perspektiven unterscheidet,
7. falsche intrinsische Kameraparameter und damit fehlerhafte bzw. ungenaue Entzerrung der Einzelbilder beider Kameras,
8. falsche extrinsische Kameraparameter und damit fehlerhafte bzw. ungenaue für die Rektifizierung benötigte Objektivausrichtung beider Kameras des Stereokamerasystems zueinander,
9. zu geringe Auflösung oder Schärfe des Aufnahmesystems,
10. Ungenauigkeit bzw. Unschärfe durch zur Stabilisierung herangezogenen größeren Vergleichsbereich,
11. zu geringe Bittiefe zur Speicherung der Disparität,
12. Vergleichsfenster bei Verfahren der Klasse I sparen je nach Größe Bereiche am Rand aus.

Die Fehler lassen sich folgenden Fehlerklassen zuordnen: keine gefundene, ungenaue und fehlerhafte Punktkorrespondenzen.

Bei ungenauen Punktkorrespondenzen, die einen Fehler von $\Delta\delta$ besitzen, ergibt sich ein Fehler in der Bestimmung der Entfernung von [490]

$$\Delta z'_0 = \frac{bf'}{\delta - \Delta\delta} - \frac{bf'}{\delta} = \frac{bf'}{\frac{bf'}{z'_0} - \Delta\delta} - z'_0 = \frac{\Delta\delta z'^2_0}{bf' - \Delta\delta z'_0}. \quad (3.25)$$

Gewöhnlich ist

$$bf' \gg \Delta\delta z'_0.$$

Für Stereokamera Nr. 4 mit einem vom Kamerahersteller angegebenen Pixelfehler von $\Delta\delta = 0,2 \text{ px}$ ist in Kopfhöhe bei $z'_0 = 608 \text{ cm}$ (angenommene durchschnittliche Personengröße von $\bar{p}^h = 176 \text{ cm}$) diese Bedingung erfüllt:

$$983 \text{ px} \cdot 24 \text{ cm} \gg 0,2 \text{ px} \cdot 608 \text{ cm}.$$

Gleichung 3.25 kann bei dieser Annahme abgeschätzt werden mit:

$$\Delta z'_0 \approx \frac{\Delta\delta z'^2_0}{bf'}. \quad (3.26)$$

Der Fehler in der Entfernungsbestimmung hängt somit ungefähr quadratisch von der Entfernung und linear vom Fehler in der Bestimmung der Punktkorrespondenz ab [491]. Abb. 3.43 zeigt diesen Zusammenhang. Ergebnisse der Entfernungsbestimmung mittels der Disparität bei Einsatz der Stereokameras Nr. 4 stehen in Abschnitt 3.6.2.

Zur Speicherung der Disparität pro Pixel reicht ein Byte nicht aus, da sonst maximal eine Disparität von 255 zulässig wäre und dabei noch keine Subpixel gespeichert wären. Um den Fehler Nr. 11 aus obiger Liste von Seite 96 nicht zu groß werden zu lassen, werden daher für gewöhnlich 2 Bytes zur Speicherung der Disparität verwendet. Dabei werden die 16 Bits je nach Bibliothek unterschiedlich belegt. Sei m die Anzahl an Bits für die ganzzahlige Disparität und damit $16 - m$ die Anzahl an Bits

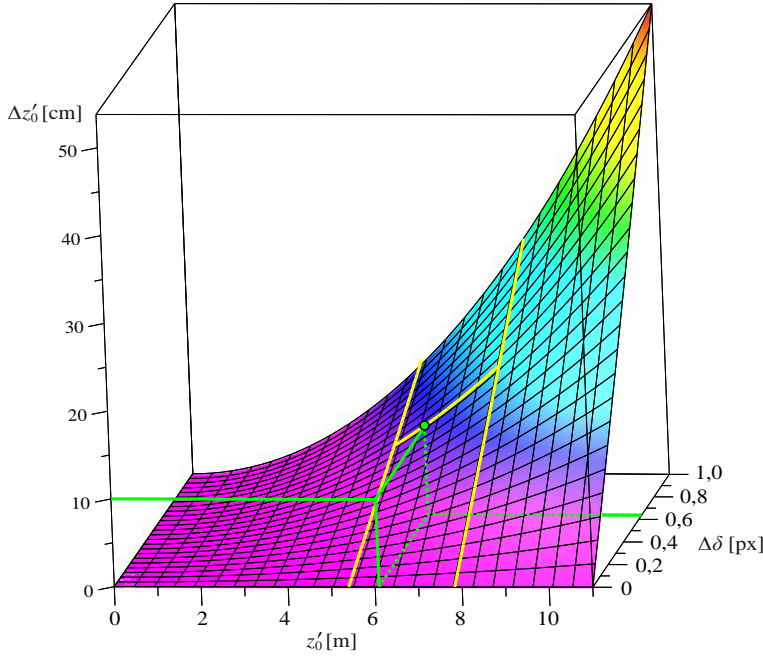


Abb. 3.43: Fehler bei der Bestimmung der Entfernung aufgrund einer fehlerhaften Disparität in Abhängigkeit zur Entfernung zur Kamera; gelbe Linien grenzen den für die ebenen Hermes-Versuche (Nr. 6) relevanten Bereich ein; grüne Linien zeigen den Punkt, der sich in den Ergebnissen der Ebenen Versuche mit dem Stereokamerasystem Nr. 4 zeigt (siehe Abschnitt 3.6.2)

zur Angabe der Subpixel, so könnte sich die Lage gleicher Punkte in beiden Perspektiven maximal um

$$2^m - 1/2^{16-m}$$

unterscheiden und ein Pixel ließe sich in $1/2^{16-m}$ Subpixel aufteilen. Die Bibliothek des Kameraherstellers von Nr. 4 hat eine Aufteilung mit $m = 8$ gewählt und beschränkt damit die maximale Disparität auf 255,996 px, was gemäß Gleichung 3.23 für Kamera Nr. 4 einer minimalen Entfernung von ca. 92 cm entspricht. Hingegen schränkt die ebenfalls in PeTrack verwendete Bibliothek OpenCV [446, 447] die Subpixel-Genauigkeit mit $m = 12$ auf $1/16$ eines Pixels ein. Daraus ergibt sich für die ebenen Versuche aus Versuchsreihe Nr. 6 eine nicht unterscheidbare Tiefe in durchschnittlicher Personengröße gemäß Gleichung 3.25 von $\Delta z'_0 \approx 0,5$ cm und somit ein maximaler Fehler von $\Delta z'_0/2 \approx 0,25$ cm.

3.5 Detektion ohne Marker

In diesem Abschnitt soll ein vom Autor entwickeltes Verfahren vorgestellt werden, das die Personenerkennung ohne Zuhilfenahme einer Markierung durchführt und so die Möglichkeit bietet, Feldstudien (Beobachtungen im natürlichen Umfeld ohne jegliche Beeinflussung) durchzuführen oder aber auch ohne Präparationsaufwand Menschenansammlungen (z. B. Versammlungen, Vorlesungen) zu nutzen,

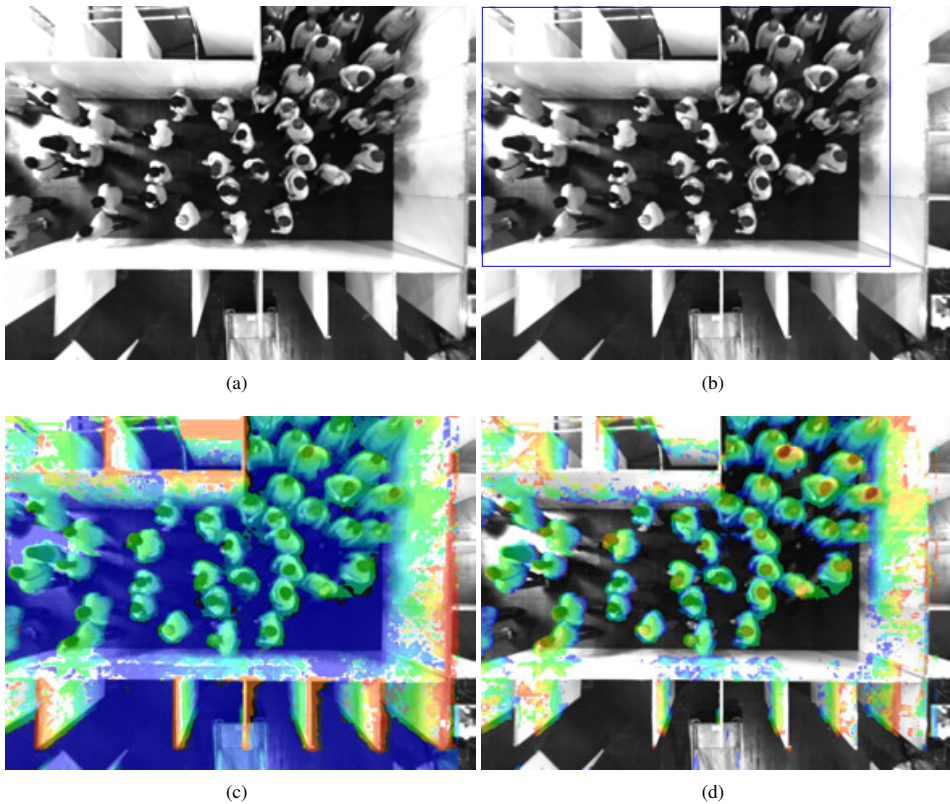
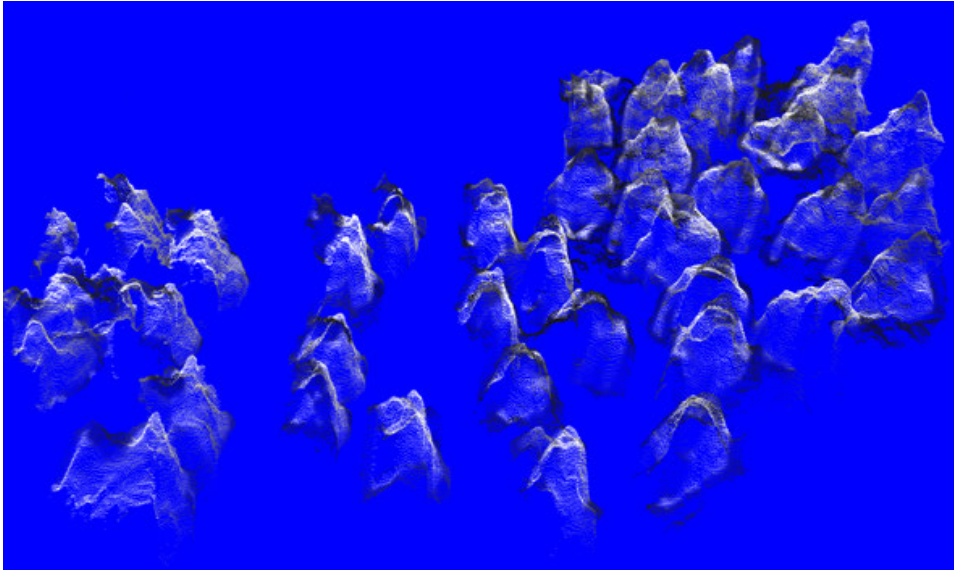


Abb. 3.44: Experiment an einer 3 m breiten Ecke aus Versuchsreihe Nr. 6, bei dem die Probanden keinen Marker trugen: (a) Bild des linken Stereosensors; (b) Bild des rechten Stereosensors mit blauem Rechteck, der den in Abb. 3.50 verwendeten Bildausschnitt markiert; (c) überlagerte farbkodierte Disparität beider Stereobilder entsprechend der Entfernung zur Kamera zwischen 534 cm (rot) und 784 cm (blau); für farblose Pixel konnte keine Disparität bestimmt werden (d) Disparität eingeschränkt auf den Bereich, der als Vordergrund für die Erkennung einer Person herangezogen wird; Entfernung zur Kamera liegt zwischen 570 cm (rot) und 735 cm (blau); Rückprojektion in das Weltkoordinatensystem in Abb. 3.45

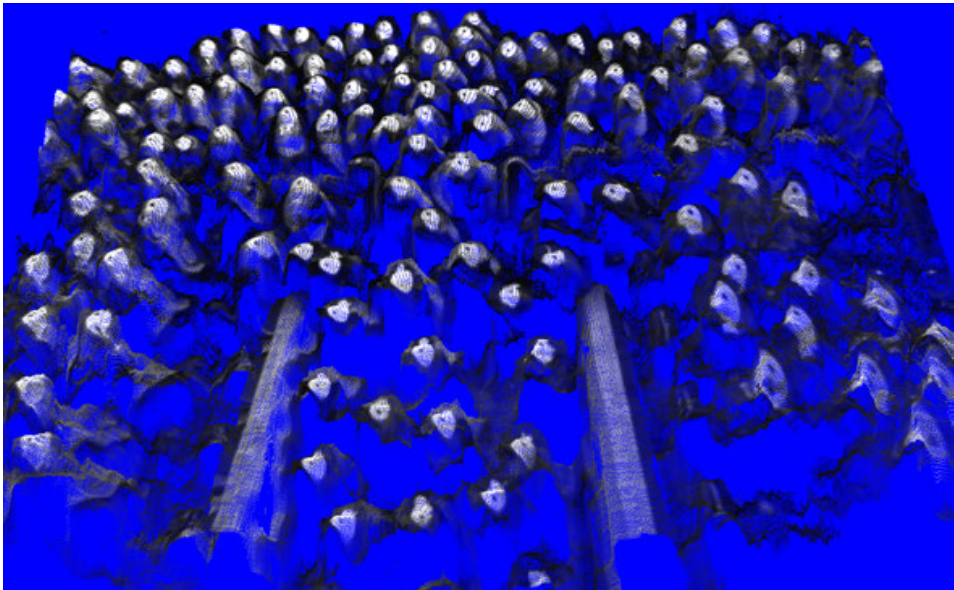
um Datenerhebungen im Feldexperiment (Beobachtung im natürlichen Umfeld mit Beeinflussung interessierender Parameter) zu realisieren. In Kapitel 4 wird näher darauf eingegangen, warum Laborversuche vorzuziehen sind, aber auch warum nicht auf Feldstudien verzichtet werden kann.

Zudem erlaubt dieses Verfahren die Güte der Personenerkennung mit Markern zu steigern (siehe Tab. 3.8).

Das markerlose Verfahren basiert auf der Formerkennung des Oberkörpers einer Person. Die Form wird direkt aus dem Disparitätsbild einer Stereokamera gewonnen. Dabei wird für eine Bewertung darauf verzichtet, die perspektivisch bestimmte Disparität durch Rückprojektion in das Weltkoordinatensystem (*plan-view statistics*) [295] zu überführen. Rückprojizierte Höhenfelder, berechnet gemäß Gleichung 3.18 und 3.24, sind als Punktwolke in Abb. 3.45 dargestellt und zeigen die Ausformung der Personen. Das Höhenfeld in Abb. 3.45 a entspricht der Disparitätsmatrix aus Abb. 3.44 d,



(a)



(b)

Abb. 3.45: Höhenfeld des Disparitätsbildes als Rückprojektion in das Weltkoordinatensystem wie es nicht für das hier vorgestellte markerlose Verfahren verwendet wird; das Höhenfeld zeigt die Ausformung der Personen; die Punkte der Punktwolke sind entsprechend der zugehörigen Pixelfarbe eingefärbt; (a) entspricht der Disparitätsmatrix aus Abb. 3.44 d; (b) entstammt dem gleichen Versuch wie Abb. 3.3 und Abb. 3.70 folgende und zeigt aufgrund der Marker besser als in a die Lage der Köpfe im Höhenfeld

eingeschränkt auf den Vordergrund, dessen Bestimmung im nächsten Abschnitt beschrieben wird. Abb. 3.45 b ist ein Höhenfeld eines Versuchs, bei dem die Personen Marker trugen, so dass die Lokalität des Kopfes im Höhenfeld besser zu erkennen ist. Dabei ist die aus einer Disparitätsmatrix pixelweise bestimmte Punktwolke im klassischen Sinne kein reines Höhenfeld, da der Verlauf der das Höhenfeld beschreibenden Fläche nicht injektiv in Bezug auf die Grundfläche ist. Das perspektivische Höhenfeld mit den Pixeln als Definitionsmenge ist hingegen injektiv.

Für die endgültige Erkennung sind einige aufeinanderfolgende Schritte nötig, die im Folgenden beschrieben werden und den Abbildungen 3.50 und 3.51 zu entnehmen sind. Abb. 3.44 zeigt das für diese Erläuterung herangezogene Experiment. Video Nr. 34 zeigt ebenfalls die Schritte der markerlosen Erkennung während eines kompletten Versuchsdurchgangs.

3.5.1 Hintergrundsubtraktion



Abb. 3.46: Personendetektion auf Basis des Höhenfeldes ohne Nutzung der Marker; keine Hintergrundsubtraktion führt für die Leitkegel zu falsch positiven Ergebnissen; die roten Zahlen geben die ermittelte Objekthöhe an

Für monokulare Aufnahmen wird ein in PeTrack integriertes Verfahren zur Hintergrundsubtraktion (*background subtraction*) bereits in Abschnitt 3.2 vorgestellt. Für Stereoaufnahmen wurde ein direkt auf den Disparitätswerten der Hintergrundszene aufsetzendes Verfahren entwickelt, das auch bei dichten langsamen Personenströmen eingesetzt werden kann.

Um Personen nur aufgrund der Oberkörperform zu erkennen, ist die Bestimmung des Vordergrundes zu empfehlen, um die Anzahl der falsch positiven Ergebnisse zu verringern. Nur im Vordergrund können sich Personen befinden. Es ist in PeTrack trotzdem möglich, die markerlose Erkennung ohne Hintergrundsubtraktion durchzuführen, was zu mehr falsch positiven Detektionen führen kann. Abb. 3.46 zeigt eine Personenerkennung ohne Marker und ohne Hintergrundsubtraktion, wodurch die Leitkegel als Personen detektiert werden. Für diesen speziellen Fall würde die Einführung einer minimalen Personengröße die Leitkegel als Personen ausschließen lassen, aber andere Objekte im Hintergrund könnten hierdurch ggf. nicht erfasst werden.

Zur Selektion des Vordergrundes wird der Bereich des Bildes als Hintergrund maskiert, der entsprechend des perspektivischen Höhenfeldes des Hintergrundes $H_{bg} = (h_{bg,uv})$ sich kaum von diesem abhebt. Die Elemente des perspektivischen Höhenfeldes $H = (h_{uv})$ sind dabei für ein Pixel mit der Koordinate $\vec{u} = (u \ v)^T$ und der Aufhängenhöhe der Kamera von a definiert als

$$h_{uv} = a - t_{uv}.$$

Das perspektivische Tiefenfeld $T = (t_{uv})$ ergibt sich dabei aus dem Disparitätsbild bzw. der Disparitätsmatrix $T' = (t'_{uv})$ mit $t'_{uv} = \delta(u)$ in Pixelzeile v gemäß Gleichung 3.24:

$$t_{uv} = \frac{sb f}{t'_{uv}}.$$

Als Hintergrund werden nun alle Pixel \vec{u} maskiert, für die gilt:

$$h_{uv} - h_{bg,uv} < \Delta h. \quad (3.27)$$

Dabei ist H_{bg} das Höhenfeld des reinen Hintergrundes, das dann erstellt werden kann, wenn sich keine Personen in der Versuchsanlage befinden.

Für den Hintergrundabstand Δh hat sich als guter Kompromiss 40 cm herausgestellt. Δh darf nicht zu klein gewählt werden, damit Ungenauigkeiten in der Bestimmung der Disparität den Vordergrund nicht unnötig erweitern. Zudem wird für die folgende Personensuche nur die Form des Oberkörpers analysiert, der sich bei nahezu allen Personen oberhalb von 40 cm befindet. Die Höhendifferenz zum Höhenfeld des Hintergrundes darf aber auch nicht zu groß gewählt werden, da aufgrund der direkten Nutzung der Disparität ohne Rückprojektion in das Weltkoordinatensystem Personen in der Nähe des Hintergrundes (z. B. Wänden) maskiert würden. Es wird nicht der Abstand zum senkrecht unter einer Person befindlichen Untergrundes als Maß herangezogen, sondern der Abstand zum Hintergrund der auf dem Sehstrahl durch das entsprechende Pixel verläuft. Für Personen mit einer Schulterbreite von $w_s = 48$ cm (Medianwert der Schulterbreite erwachsener Männer [492, 493]), die ohne Seitenabstand parallel zu einer Wand gehen (Normalfall bei sich bewegenden Personen), ergibt sich ein maximaler Öffnungswinkel 2α von

$$2\alpha = \arcsin \frac{w_s/2}{\Delta h} \approx 74^\circ,$$

um die Kopfmitte im Vordergrund zu belassen (siehe Abb. 3.47). Dies ist auch für die verwendete weitwinklige Stereokamera Nr. 4 (siehe Tab. 4.2) ausreichend. In Abb. 3.48 ist die Erosion des Vordergrundes im Bereich der Schulter der oberen Person zu erkennen, die durch die Nähe zur Wand dem Hintergrund zugeschrieben wird. Die Pixel des Hintergrundes sind ausgegraut.

Für ein möglichst gleichmäßiges Höhenfeld wird die Disparitätsmatrix mit einer großen Maskengröße erstellt und Bereiche fehlender Werte in der Matrix aufgefüllt, indem diese durch lineare Interpolation innerhalb einer Pixelzeile aus den nächstgelegenen Werten bestimmt werden. Am Bildrand wird der dem Rand am nächsten gelegene Wert fortgeschrieben. Diese Interpolation wird für die Disparitätsmatrix des Hintergrundes einmalig wie auch für jedes Disparitätsbild der Bildsequenz des Versuches durchgeführt, wobei bei der Interpolation innerhalb der Disparitäten der Sequenz nur kleine Lücken von maximal 15 Pixel geschlossen werden. Wie in Abschnitt 3.4.2 beschrieben, kommt es zu fehlenden Disparitäten, wenn z. B. aufgrund fehlender Textur keine korrespondierenden Pixel gefunden werden können oder aber Bereiche nicht von beiden Kameraperspektiven aus einsehbar sind (siehe z. B. Überlappungsbereiche der Kameras für den Kopf in Abb. 3.35). Zur Erzeugung der Disparitätsmatrix in Abb. 3.50 wurde das SGBM-Verfahren mit einer maximalen Maskengröße von 11 px verwendet.

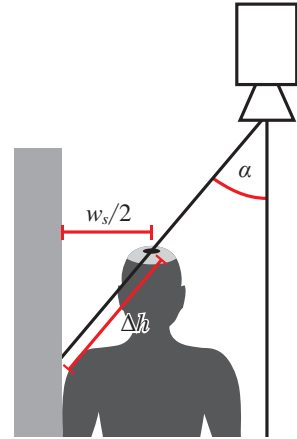


Abb. 3.47: Für Personen mit Schulterbreite w_s und einer Kamera mit maximalem Öffnungswinkel 2α darf zur Erhaltung der Kopfmitte im Vordergrund der Hintergrundabstand maximal Δh betragen

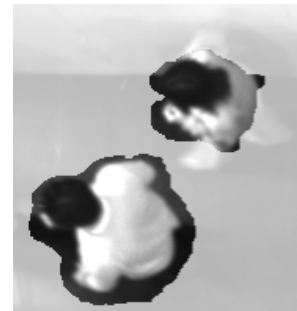


Abb. 3.48: Vordergrunderosion durch perspektivisches Höhenfeld; ausgegrauter Hintergrund bedeckt Schulter der oberen Person nahe der Wand

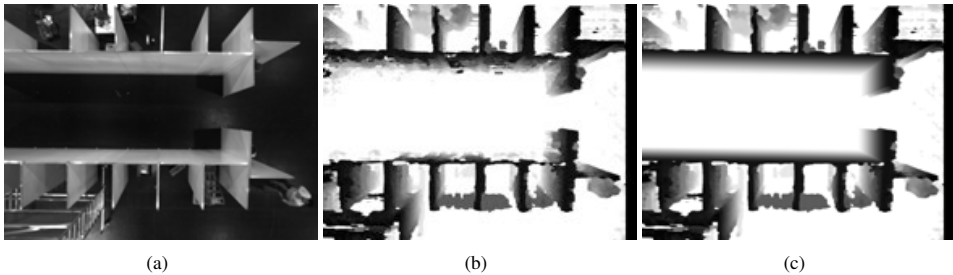


Abb. 3.49: (a) Kamerabild des Hintergrundes; (b) Disparitätsmatrix mit vielen Fehlern aufgrund geringer oder wiederholender Bildstruktur; (c) manuell angepasste Disparitätsmatrix im Bereich, in dem Personen erkannt werden sollen

Trotzdem kann das Resultat z. B. für strukturlose Hintergründe unbefriedigende Ergebnisse liefern (siehe Abb. 3.49 b). Aus diesem Grund gibt es für diese Fälle in `PeTrack` die Möglichkeit, das berechnete Höhenfeld des Hintergrundes zu exportieren, manuell nachzuarbeiten (siehe Abb. 3.49 c) und wieder zu importieren.

Statt einer manuellen Anpassung könnte auch ein träge angepasstes Minimum des Höhenfeldes eine Verbesserung bewirken. Sei $H_i = (h_{i,uv})$ das Höhenfeld des Bildes B_i einer Bildfolge B , so ergeben sich damit die Elemente des adaptierten Höhenfeldes des Hintergrundes pixelweise mit

$$h_{bg,uv} \approx \min_i h_{i,uv}.$$

Eine zusammenhängende Region im Vordergrund, deren Fläche kleiner ist als die Fläche von $15\text{ cm} \times 20\text{ cm} = 300\text{ cm}^2$, die ein Kopf in minimaler Personengröße bedecken würde, wird ebenfalls als Hintergrund maskiert, da sich in diesem Bereich keine Person befinden kann. Dadurch wird ebenfalls das Rauschen und die Anzahl der im Folgenden zu untersuchenden Bereiche verringert. Bei markerlosen Versuchen auf unebenen Untergründen (z. B. Treppen) müsste die maximale Flächengröße, die zum Ausschluss aus dem Vordergrund führt, an die umliegenden Disparitätswerte angepasst werden, da die Größe der von einer Person bedeckten Fläche sich je nach Abstand zur Kamera ändert. Auch werden Bereiche des Hintergrundes, die kleiner als 120 cm^2 sind, dem Vordergrund zugeschrieben, wenn sie vom Vordergrund umschlossen sind, um kleine Löcher im Vordergrund zu entfernen. Am Ende wird ähnlich der morphologischen Basis-Operation des Öffnens und Schließens bei der Bestimmung der Position des Farbmarkers (siehe Abb. 3.15) auf die Maske des Vordergrundes eine den Rand glättende Erosion und zweifache Dilatation mit einem 3×3 -Kern (*kernel*) angewendet.

3.5.2 Isolinien gleicher Disparität

Die Formerkennung einer Person zur Positionsbestimmung wird direkt im Disparitätsbild bzw. im davon abgeleiteten perspektivischen Höhenfeld durchgeführt. Daher ist die Güte des Disparitätsbildes entscheidend für das Auffinden der Höhenstruktur einer Person. Die Wahl des Verfahrens zur Bestimmung der Disparitäten kann Abschnitt 3.6.5 entnommen werden. Im Folgenden dieses Abschnittes wird daher, wenn nichts anderes gesagt wird, das dort für die markerlose Detektion empfohlene SGBM-Verfahren mit einer Vergleichsfenstergröße von 11 px verwendet. Das abgeleitete Höhenfeld wird perspektivisches Höhenfeld genannt, da es nicht wie bei einem normalen zweidimensionalen

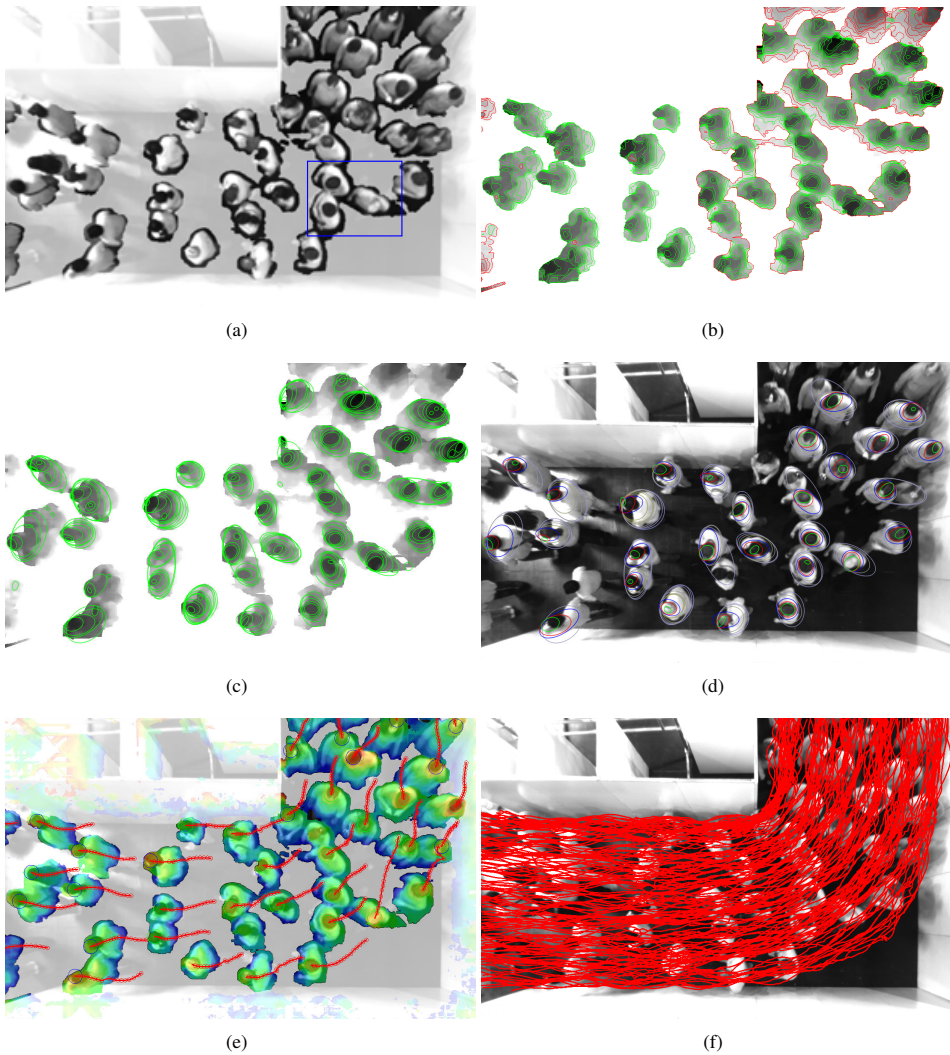


Abb. 3.50: Schritte der markerlosen Personendetektion am Beispiel des Versuches aus Abb. 3.44; (a) ausgegrauter Hintergrund zeigt den für die weitere Verarbeitung berücksichtigten Vordergrund; blaues Rechteck umrandet den Ausschnitt, der in Abb. 3.51 vergrößert dargestellt ist; (b) Isolinien gleicher Disparität in Abständen von 10 cm im Bereich des Vordergrundes, aufgetragen auf das als Grauwertbild dargestellte Disparitätsbild; grüne Linien umkreisen Bereiche, die näher an der Kamera liegen, rote Linien entsprechend entferntere Bereiche; (c) die grünen Isolinien approximierende Ellipsen; (d) Ellipsen gruppiert nach sich einschließenden Isolinien zur Erstellung der die Personen identifizierende Ellipsenpyramiden; grüne Ellipsen sind innerhalb der Pyramide der Kamera am nächsten, gefolgt von den roten und blauen Ellipsen; (e) rote Linien sind Laufwege der Personen in der letzten Sekunde; der Mittelpunkt des kopfgroßen grünen Kreises zeigt die aktuelle Position, der der roten Kreise die Position in den vorangegangenen 16 Frames an; (f) alle 162 Laufwege des Versuches

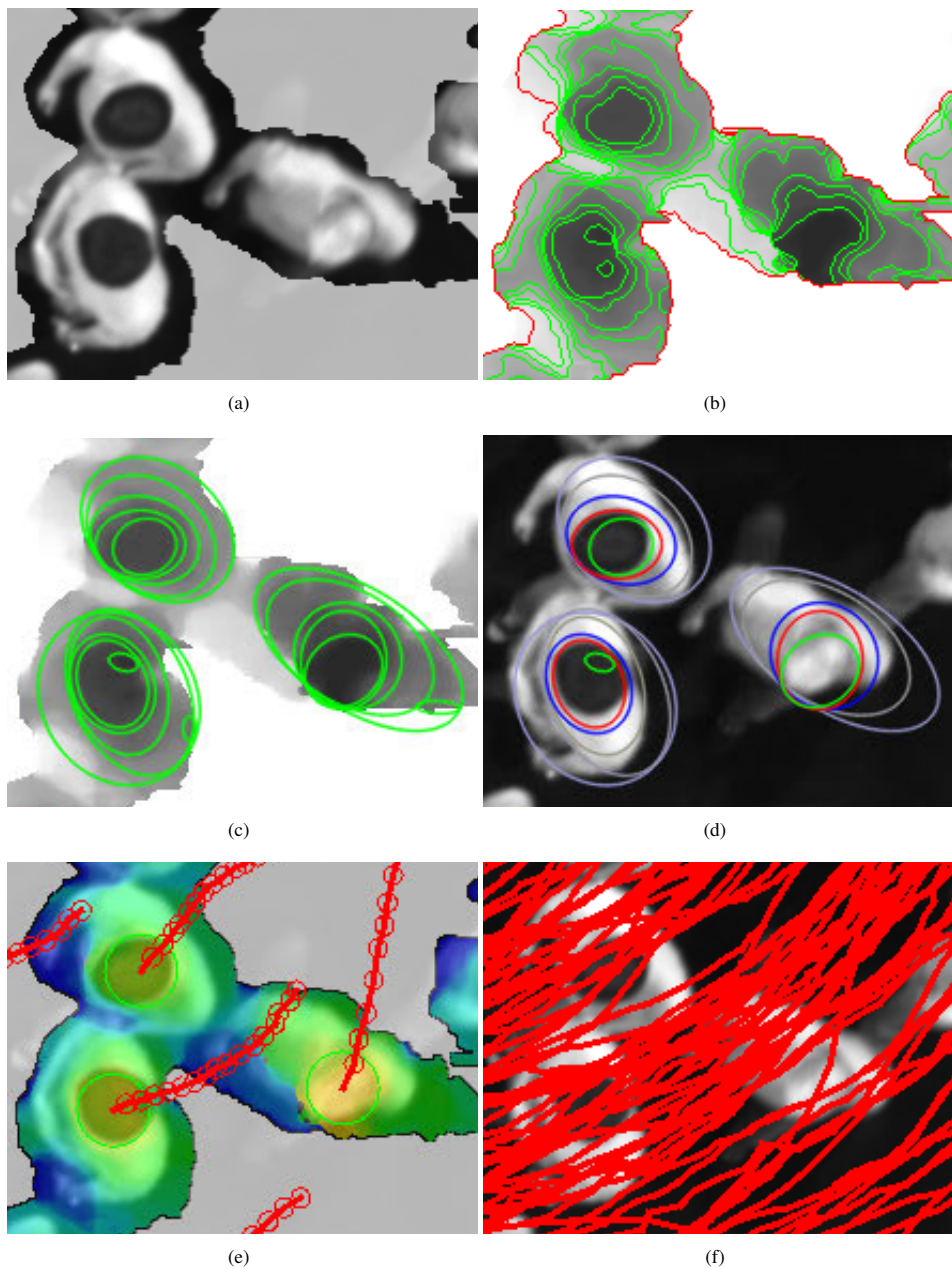


Abb. 3.51: Ausschnitte der Schritte der markerlosen Personendetektion aus Abb. 3.50

skalaren Höhenfeld zur Beschreibung eines Höhenreliefs äquidistant angeordnete Höhenwerte enthält, sondern jedes Element des Feldes die Höhe des Objektes auf dem Bildpunkt enthält (siehe Abb. 3.52).

Fehlende Disparitätswerte innerhalb einer Zeile werden wie bereits erwähnt interpoliert. Die Wirkung der Interpolation ist in Abb. 3.51 zu sehen. Die farbkodierten Disparitätswerte, die das Verfahren direkt berechnet, sind in e halbtransparent auf das Bild gelegt, im entsprechenden Grauwertbild von b und c sind die kleinen Lücken im Disparitätsbild von e u. a. vor dem Kopf der rechten und oberen Person durch zeilenweise Interpolation gefüllt worden.

Größeren Effekt hat die Interpolation bei Verfahren oder kleineren Maskengrößen, die mehr fehlende Disparitätswerte aufweisen. Zur Demonstration ist zur Erzeugung der Disparitätsmatrix in Abb. 3.53 das BM-Verfahren mit einer Maskengröße von 17 px verwandt worden. Dabei entstehen mehr Lücken in der Disparitätsmatrix, die als weiße, dem Hintergrund zugesprochene Bereiche erkennbar sind. In Abb. 3.53 a ist keine Interpolation durchgeführt worden, wodurch bei der oberen Person rot umrandete Lücken bleiben, was bei der rechten Person sogar dazu führt, dass die rechte Schulter eine getrennte Erhöhung darstellt und ggf. in Folgeschritten als separate Person erkannt würde. Das unruhigere Höhenfeld in Abb. 3.53 aufgrund der geringeren Maskengröße ist gut an den zackigeren Linienverläufen im Vergleich zu Abb. 3.54 zu erkennen.

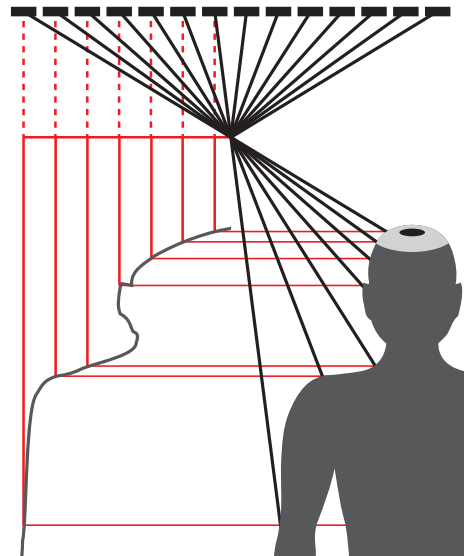


Abb. 3.52: Perspektivisches Höhenfeld als dunkelgraue Kontur für Pixel, dessen Entfernung zur Kamera an rechter Person abgenommen wurden

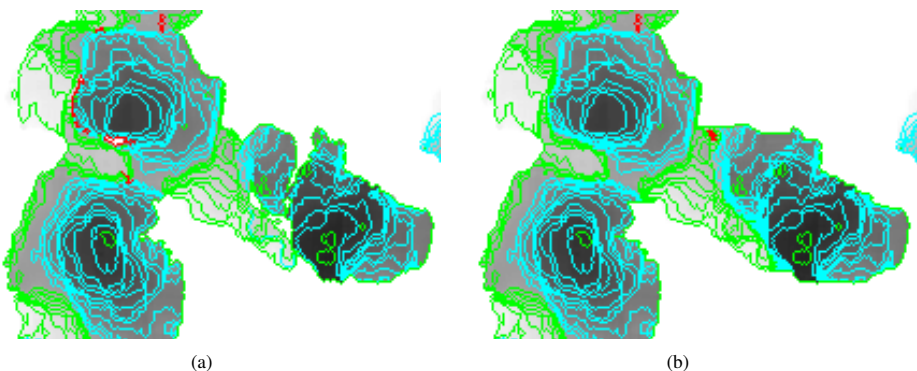


Abb. 3.53: Grauwertbild der Disparitätsmatrix ermittelt mit dem BM-Verfahren und einer Maskengröße von 17 px; wie in Abb. 3.54 gefärbte Isolinen in 5-cm-Schritten; Linien der Farbe Cyan umschreiben wie die grünen Linien erhöhte Bereiche, werden aber entgegen der grünen Linien an nachfolgende Schritte weitergeleitet; der Ausschnitt entspricht dem in Abb. 3.51; (a) ohne Interpolation; (b) mit zeilenweiser Interpolation kleiner Lücken

Die als Hintergrund maskierten Pixel werden in der Disparitätsmatrix auf den kleinsten Disparitätswert des Vordergrundes gesetzt.

Innerhalb des so angepassten Disparitätsbildes bzw. entsprechenden Höhenfeldes werden nun gerichtete Isolinien gleicher Disparität bzw. Höhenlinien gleicher Entfernung zur Kamera in äquidistanten Entfernungsschritten bestimmt. Dabei wird das gesamte Größenspektrum des Vordergrundes abgetastet, um sowohl sehr große Personen als auch Kinder auffinden zu können. Die Nutzung von großen Masken bei der Disparitätsbestimmung, der Interpolation fehlender Disparitätswerte und dem Setzen des Hintergrundbereiches auf die größte Entfernung zur Kamera führt zu besonders formentsprechenden Linienverläufen der Isolinien. Abb. 3.50 b bzw. 3.51 b zeigt diese Verläufe. Dabei umschreiben grüne Linien höherliegende Bereiche bzw. Gebiete größerer Disparitäten und rote Linien entsprechend entferntere Bereiche bzw. Gebiete kleinerer Disparitäten. Diese entfernteren Bereiche entstehen insbesondere bei Höhenlinien, die bis zum Bildrand gehen, da dieser als offen angesehen wird, und werden im Folgenden nicht weiter betrachtet. Eine Darstellung der Disparitätsmatrix als Grauwertbild ist gewählt worden, damit die farbigen Isolinien besser zu erkennen sind.

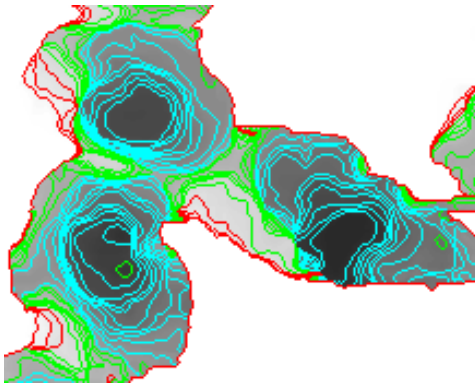


Abb. 3.54: Abb. 3.51 b mit Isolinien im Abstand von 5 cm; nur Linien in der Farbe Cyan werden an nachfolgende Schritte weitergeleitet

Neben Linien, die entfernte Bereiche umschließen, werden auch solche Isolinien verworfen und nicht mehr an die folgenden Schritte weitergereicht, die im Vordergrund Bereiche einschließen, die in minimaler Entfernung zur Kamera (in ebenen Versuchen entspricht dies der maximalen Kopfhöhe) kleiner als 20 cm^2 oder in maximaler Entfernung zur Kamera (in ebenen Versuchen entspricht dies 40 cm über Bodenhöhe) größer als 4500 cm^2 sind. Damit werden zum einen kleine Erhebungen bzw. Ungenauigkeiten in der Disparitätsmatrix beseitigt, auch wenn dadurch ggf. die höchste Erhebung auf dem Kopf unberücksichtigt bleibt. Zum anderen werden große Flächen, die nicht nur eine Person umschreiben können, verworfen und damit unnötige Berechnungen in Folgeschritten vermieden. Nur die Cyan eingefärbten Linien in

Abb. 3.54 werden den nachfolgenden Schritten weitergeleitet.

Eine weitere sinnvolle Forderung scheint ein möglichst kleines ausdehnungs bereinigtes Verhältnis der Länge einer Isolinie zu der Größe ihrer umschlossenen Fläche (minimal $2/r$ bei einem Kreis vom Radius r) zu sein, um Isolinien mit starken Ausbuchtungen zu verwerfen. Dies führte jedoch zu keinen verbesserten Ergebnissen. Zudem würde daraus resultieren, dass eine Folge sich einschließender Isolinien entlang der Höhenstufen nicht mehr zusammenhängend vorliegen muss, sondern innerhalb solcher Folgen Höhen ohne Isolinien existieren könnten. Daher wurde diese Forderung nicht in das Verfahren integriert.

Durch das Setzen der Disparität des Hintergrundes auf den kleinsten vorkommenden Disparitätswert und aufgrund der perspektivischen Sicht, ist der Verlauf der Disparität bei Erhöhungen insbesondere auf der der optischen Achse abgewandten Seite nicht stetig und fällt abrupt ab, so dass dort Isolinien unterschiedlicher Höhen nahezu übereinander liegen.

Die äquidistante Schrittweite, in der Isolinien im Disparitätsbild gesucht werden, wurde zwischen

5 cm und 10 cm variiert. Da am Ende nur wenige Isolinien bzw. deren abgeleitete Formen eingehender untersucht werden (siehe Abschnitt 3.5.4), spielt die Wahl der Schrittweite keine signifikante Rolle, sollte aber maximal 10 cm betragen, um relevante Formfaktoren testen zu können. Zum Beispiel zeigt Abb. 3.50 Bilder des Verfahrens mit einer Schrittweite von 10 cm und Abb. 3.84 mit einer Schrittweite von 5 cm. In Abb. 3.42 d ist zu sehen, dass bei der verwendeten Basislänge, dem verwendeten Abstand zwischen Kamera und Kopf der Versuchsteilnehmer und dem interessierenden Entfernungintervall der Zusammenhang zwischen Disparität und Kameraentfernung eines Pixels nahezu linear ist, so dass auf eine Umrechnung der Disparitätsmatrix zu einem perspektivischen Höhenfeld mit nur geringem Genauigkeitsverlust verzichtet und die Disparitätsmatrix direkt äquidistant abgesprochen werden könnte.

3.5.3 Approximierende Ellipsen

Die gemäß vorherigem Abschnitt ermittelten Höhenlinien werden wie bei der Suche nach Markern mit Strukturelementen in Abschnitt 3.2.1 durch Ellipsen approximiert, da sie der Körperform nahe kommen und einen einfacheren Zugriff auf die Form (u. a. Seitenverhältnis), Orientierung, Größe und Schwerpunkt bzw. Position der Isolinien erlauben. Auch hier wird wieder mittels der kleinsten Fehlerquadratsumme der Abstand zu den Punkten jeder Isolinie minimiert [464].

Einige Isolinien in Abb. 3.50 b bzw. 3.51 b wurden nicht durch Ellipsen approximiert, da, wie im vorherigen Abschnitt beschrieben, Isolinien, die sehr kleine und große Bereiche umschließen, verworfen werden, so dass in Abb. 3.50 c bzw. 3.51 c die Anzahl der approximierenden grünen Ellipsen kleiner als die der Isolinien ist (siehe Abb. 3.54).

Zudem werden nach der Approximation Ellipsen verworfen, deren numerische Exzentrizität größer als $\sqrt{15}/4$ bzw. das Seitenverhältnis der Halbachsen größer als 4:1 ist, um Ellipsenformen zu verwerfen, die dem horizontalen Personenquerschnitt des Oberkörpers nicht ähneln. Diese Einschränkung wird jedoch nur für Ellipsen mit einer Hauptachse gemacht, die kleiner als 120 cm ist, da große Ellipsen in dichten Menschenmengen mehrere Personen beinhalten können. Um aber eine ausreichende Zahl an Ellipsen zu haben, die jeweils die Form einer Person beschreibt, wird auf die Löschung der großen Ellipsen verzichtet. Gemäß Abb. 3.57 tauchen diese Ellipsen erst unterhalb der Schulterhöhe auf.

Alle Werte zur Auswahl der Isolinien, Ellipsen und im Folgenden bei der Erstellung der Pyramiden von Ellipsen sind heuristisch und wurden mit dem Ziel gewählt, Elemente zu verwerfen, die nicht dem Höhenrelief einer Person entsprechen können.

3.5.4 Ellipsenpyramiden

Das Höhenfeld wird schrittweise (in 5- bis 10-cm-Schritten) von oben nach unten – vom Kopf bis zu den Knien – abgetastet und Höhenlinien bzw. ihre entsprechenden Ellipsen erzeugt. Dabei liegt es in der Natur des Höhenfeldes, dass die Isolinien und damit annähernd auch die Ellipsen sich aufweiten und umschließen.

Als letzten Schritt zur Bestimmung der Position einer Person soll nun für jede Person eine Pyramide von zusammengehörigen Ellipsen bzw. ein pyramidaler Ellipsenstapel (kurz bezeichnet als PES, *Pyramidal Ellipses Stack*) aufgebaut werden, der die Körperform nachbildet (siehe Abb. 3.50 und 3.51).

Werden für eine Höhe neue Ellipsen erzeugt, können folgende Fälle für jede dieser Ellipsen auftreten (siehe Abb. 3.55); das jeweils angegebene Verhältnis ist das Verhältnis der Anzahl neuer Ellipsen zur

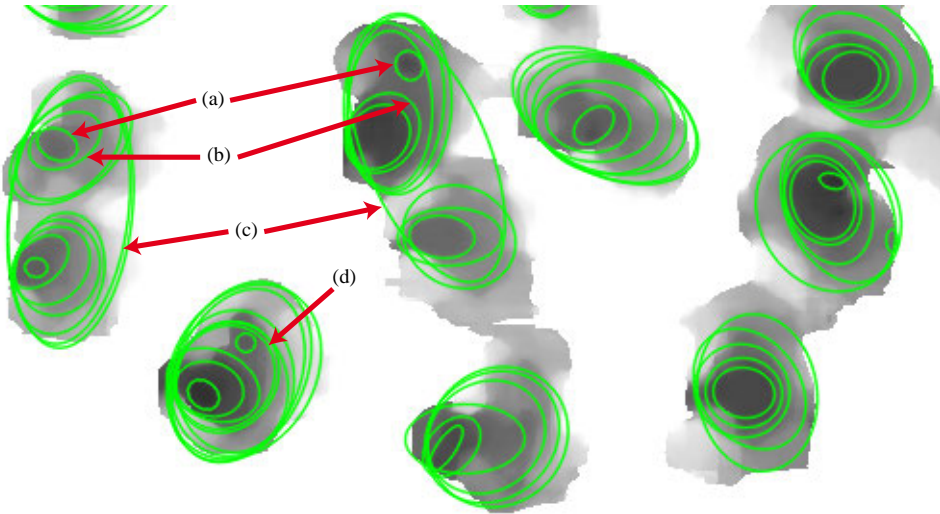


Abb. 3.55: Fälle (entsprechend der Liste auf Seite 110) bei der Zuordnung neuer Ellipsen zu bestehenden Ellipsenpyramiden; Ausschnitt aus Abb. 3.50 c; (a) Fall 1; (b) Fall 2; (c) Fall 3a mit zwei gültigen PES; (d) Fall 3a mit einem gültigen PES; Fall 4 ist Abb. 3.56 zu entnehmen

Anzahl existierender PES; die Prozentzahl gibt die Wahrscheinlichkeit des Auftretens beim Versuchsdurchlauf von Abb. 3.50 an:

1. es existiert kein vorheriger PES, dessen Mittelpunkt der Ellipse der darüberliegenden Höherebene innerhalb der neuen Ellipse liegt (1:0) (18,7%)
 ⇒ ein neuer PES wird mit der neuen Ellipse begonnen,
2. es existiert genau ein PES, dessen Mittelpunkt der Ellipse der darüberliegenden Höherebene innerhalb der neuen Ellipse liegt (1:1) (74,2%)
 ⇒ die neue Ellipse wird als unterstes Element diesem PES zugeordnet,
3. es existieren mehrere PES, deren Mittelpunkt der Ellipse der darüberliegenden Höherebene innerhalb der neuen Ellipse liegt (1:n) (7,1%)
 ⇒ je nach Beschaffenheit jedes PES:
 - (a) der PES beschreibt eine Erhebung, die höher als 10 cm ist und deren unterste Ellipse eine Hauptachse besitzt, die länger als 20 cm ist (6,5%)
 ⇒ die neue Ellipse wird als unterstes Element diesem PES zugeordnet,
 - (b) kein PES erfüllt die vorherige Bedingung (0,6%)
 ⇒ die neue Ellipse wird als unterstes Element dem PES zugeordnet, dessen Mittelpunkt der untersten Ellipse dem Mittelpunkt der neuen Ellipse am nächsten ist,
4. der Mittelpunkt der untersten Ellipse eines PES liegt in mehreren neuen Ellipsen (n:1) (0,02%)
 ⇒ nur die Ellipse wird dem PES zugeordnet, deren Mittelpunkt dem Mittelpunkt der untersten Ellipse des PES am nächsten liegt

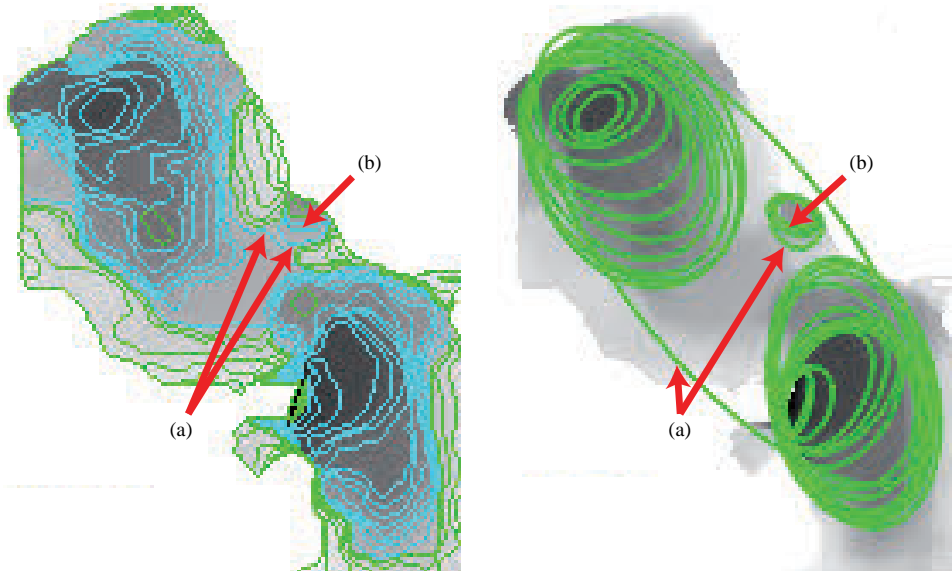


Abb. 3.56: Korrespondierende Ellipsen neuer Höhenlinien einer Höhe (a) umschließen beide PES (b); gemäß Fall 4 wird die kleinere Ellipse diesem PES zugeordnet

Das Verhältnis ($n:m$) mit $n, m > 1$ muss nicht gesondert betrachtet werden, da dies durch die sukzessive Anwendung des Falles 4 auf den Fall 3 automatisch abgearbeitet wird.

Die Einschränkung in Fall 3a soll bewirken, dass leichte Erhebungen z. B. auf der Schulter (siehe Abb. 3.55 d) keine weitere Person erzeugen sowie Erhebungen, bevor sie eine gemeinsame umschließende Ellipse haben dürfen, eine Mindestgröße besitzen müssen, die der Kopfgröße im Höhenfeld entspricht. Eine reale Kopflänge von 20 cm führt im geglätteten Höhenfeld in mittlerer Kopfhöhe zu einer Hauptachsenlänge von über 20 cm. Hierbei ist zu beachten, dass die glättende Eigenschaft der großen Vergleichsfenstergrößen der Verfahren zur Bestimmung der Disparitäten auf Pixelebene glätten, so dass diese Eigenschaft bei entfernteren Objekten einen größeren Einfluss auf die metrische Ausweitung hat als bei nahen Objekten. Die 20 cm sind so gewählt, dass die Fallunterscheidung auch für Personen korrekt verläuft, die sich beliebig nah an der Kamera befinden.

Fall 4 tritt mit 0,02% sehr selten auf. Abb. 3.56 zeigt solch eine seltene Konstellation. Die Isolinien bzw. korrespondierenden Ellipsen von a werden auf gleicher Höhe im Höhenfeld erzeugt und sollen den bestehenden PES zugeordnet werden. Dabei liegt der PES mit dem Mittelpunkt der untersten Ellipse b innerhalb beider neuen Ellipsen. Gemäß der Fallbeschreibung wird die kleinere neue Ellipse dem bei b beginnenden PES zugewiesen, da die Mittelpunkte sich näher sind. Der Grund für dieses Höhenfeld ist ein über der Schulter getragener Rucksack, der sich im Höhenfeld zwischen zwei nah beieinander laufenden Personen befindet.

3.5.5 Identifikation

Nach der Erstellung von PES-Kandidaten für Körperformen von Personen soll nun überprüft werden, ob die PES die Höhenstruktur eines Menschen nachbilden. Eine Vielzahl von Ansätzen zur Auswahl

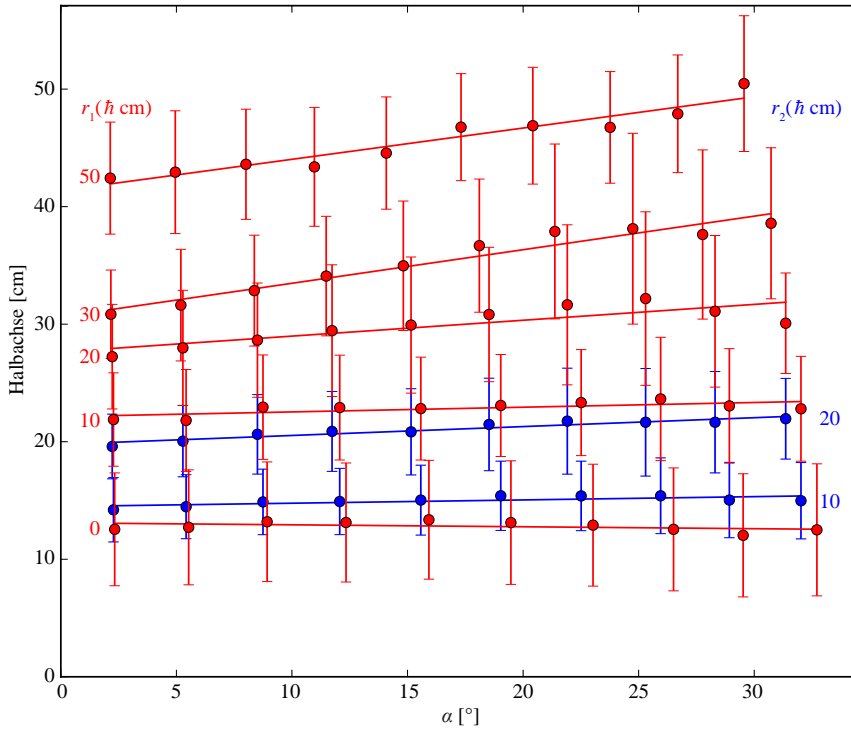


Abb. 3.57: Auswahl von gruppierten Längen der Halbachsen für die für das Verfahren relevanten Ellipsen der Ebenen PES(0 cm), PES(10 cm), PES(20 cm), PES(30 cm) und PES(50 cm); die Schnitte der Höhenlinien liegen aufgrund der Schrittweite von 5 cm im Mittel bei 2,5 cm, 12,5 cm, 22,5 cm, 32,5 cm und 52,5 cm; Fehlerbalken zeigen die einfache Standardabweichung, in deren Bereich ca. 68% der Radien liegen

der richtigen PES, die eine Person formen, sind denkbar (u. a. Mustererkennung, künstliche neuronale Netze). Einige Eigenschaften der PES werden im Folgenden untersucht und könnten alle zur Auswahl herangezogen werden. Nach ausführlicher Analyse gültiger eine Person beschreibende PES wurden folgende einfach zu testende, aber trotzdem effektive, Ausschlusskriterien entwickelt, die auf der menschlichen Wahrnehmung beruhen.

PES(\bar{h}) sei die Ellipse innerhalb des PES, die der Höhenlinie mit einer Entfernung von \bar{h} unterhalb der jeweiligen Personengröße am nächsten ist, und $r_1(\bar{h})$ und $r_2(\bar{h})$ die Längen ihrer großen bzw. kleinen Halbachse. α sei der Betrachtungswinkel zur optischen Achse in Grad (siehe Abb. 3.47). PES werden zur Sicherstellung von menschenähnlichen Formfaktoren verworfen, wenn

1. die Höhe geringer als 20 cm ist,
2. die Hauptachsenlänge von PES(10 cm) den Wert von 68 cm überschreitet:

$$r_1(10\text{cm}) > 34\text{cm},$$

3. die Nebenachsenlänge von PES(20 cm) den Wert von $(22 + 0,14\alpha/^\circ)$ cm unterschreitet:

$$r_2(20\text{cm}) < (11 + 0,07\alpha/^\circ)\text{cm und}$$

4. die unterste detektierte Ellipse (\bar{h} unterhalb der Personengröße) einen Flächeninhalt von weniger als 1250 cm^2 besitzt:

$$r_1(\bar{h}) \cdot r_2(\bar{h}) \cdot \pi < 1250 \text{ cm}^2.$$

Die aufgeführten Eliminationsgründe sollen

1. eine leichte Erhöhung, wie sie auf der Schulter oder durch einen Rucksack entstehen, nicht als separate Person ausweisen; 20 cm entspricht dabei dem Höhenunterschied von der Oberseite des Kopfes bis zum Mund; ausgehend von den am Ende der Identifikation bestimmten Personen führt die Verringerung der minimalen Personengröße auf 15 cm zu 7 weniger falsch negativen und 1 zusätzlichen falsch positiven Detektionen und die Erhöhung auf 25 cm zu 21 zusätzliche falsch negativen und 2 weniger falsch positiven Detektionen (siehe Tab. 3.3); die aufgrund des verbindenden Trackings wichtigere Falsch-Ermittlungs-Rate wird durch eine Erhöhung verringert, jedoch steigt dabei die Falsch-Negativ-Rate um 50%, so dass die Wahl von 20 cm einen guten Kompromiss darstellt,
2. eine für den menschlichen Kopf zu große Ausdehnung des PES in Augenhöhe vermeiden; die Länge der großen Halbachse r_1 (10 cm) liegt durchschnittlich zwischen 21,9 cm und 23,6 cm; die Standardabweichung ist in Augenhöhe für alle Winkel am geringsten und liegt zwischen $\pm 4,0$ cm und 5,2 cm (siehe Abb. 3.57), so dass im Durchmesser von $2(23,6 + 2 \cdot 5,2) = 68$ mehr als ca. 95% der Ellipsen liegen; alle PES mit schwarzen Punkten rechts der schwarzen Linie in Abb. 3.59 bzw. mit roten Punkten oberhalb der schwarzen Linie in Abb. 3.58 werden verworfen; da $r_2 \leq r_1$, ist die Gesamtausdehnung durch diese Einschränkung begrenzt; Eliminationsgrund 3 sorgt für die Begrenzung in orthogonaler Richtung,
3. PES verwerfen, die über die gesamte Mindesthöhe von 20 cm zu klein für einen Kopf sind, wie es bei einem erhobenen Arm der Fall wäre; die Abhängigkeit vom Blickwinkel α ist gewählt worden, da der Einfluss des Winkels auf die entsprechende Nebenachse doppelt so groß wie für den vorangegangenen Fall ist (siehe Abb. 3.57); die Länge der kleinen Halbachse r_2 (20 cm) bewegt sich im Bereich von $19,6 \text{ cm} \pm 2,8 \text{ cm}$ bis $21,8 \text{ cm} \pm 4,6 \text{ cm}$; die Gerade ist so gewählt worden, dass durch den Abstand der doppelten Standardabweichung ca. 95% der Kandidaten erhalten bleiben; Abb. 3.59 zeigt an der hellgrünen Linie für einen mittleren Blickwinkel von 15° , dass die Anzahl solcher Fälle (grüne Punkte unterhalb dieser Linie) gering ist,
4. eine Mindestausdehnung des gesamten PES gewährleisten, um Objekte auszuschließen, die kleiner als eine Person sind, wobei 1250 cm^2 nur einem durchschnittlichen Durchmesser von 40 cm entspricht, um auch Personen in dichten Menschenmengen unter Anwendung eines glättenden Disparitäts-Verfahrens nur über Ihren Kopf detektieren zu können; Abb. 3.59 zeigt, dass bei minimaler Höhe des PES von 20 cm (grüne Punkte) nur wenige Ellipsen unterhalb der dunkelgrünen Kurve ($r_2(\bar{h}) = 1250/(\pi \cdot r_1(\bar{h}))$) liegen, die der Halbachsenkombination eines Flächeninhalts von 1250 cm^2 entspricht; für gewöhnlich greift dieser Test unterhalb von PES(20 cm), da ca. 90% der PES höher als 20 cm sind.

Sollten bei der benutzten Schrittweite für die Erstellung der Isolinien die aufgeführten Höhen nicht existieren, so wird die nächstgelegene Höhe gewählt. Dabei ist zu beachten, dass die oberste Ellipse eines PES die Person im Durchschnitt erst in halber Schrittweite unterhalb der Körpergröße schneidet. Für die in dieser Arbeit verwendeten Schrittweiten von 5 cm und 10 cm, wurde für den Test Nr. 2 der 3. bzw. 2. Schnitt und für den Test Nr. 3 der 5. bzw. 3. Schnitt verwendet. Eine minimale Schrittweite der Höhenlinien von 10 cm ist für die gewählten Tests ausreichend.

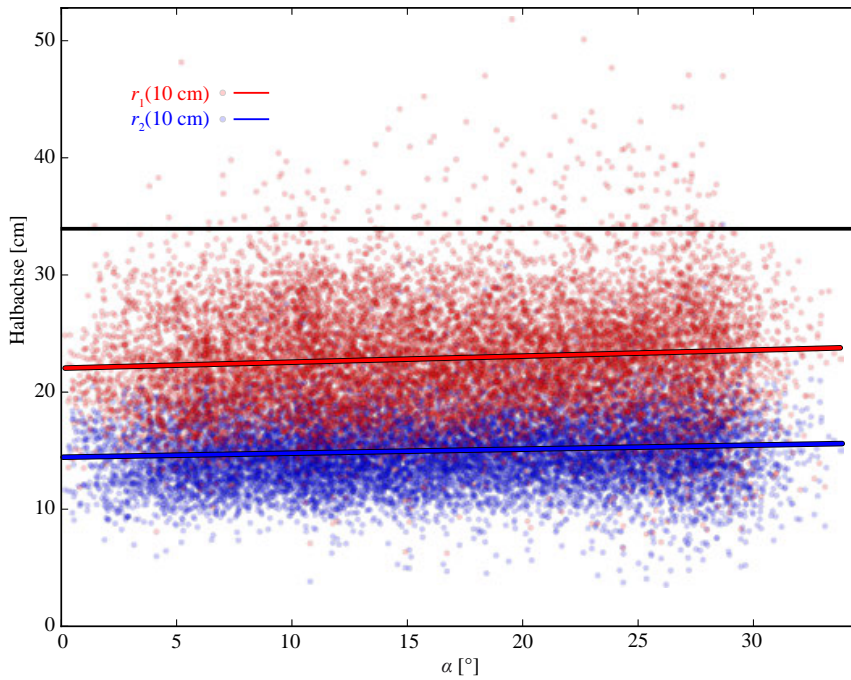


Abb. 3.58: Große (rot) und kleine (blau) Halbachsen von PES(10 cm) eines Versuches in Abhängigkeit vom Blickwinkel; die Regressionsgeraden zeigen die geringe Abhängigkeit der Halbachsen in Augenhöhe vom Winkel; die schwarze Linie bei einer Länge der Halbachse von 34 cm zeigt die Grenze zur Ablehnung eines PES aufgrund von Fall 2 auf Seite 112

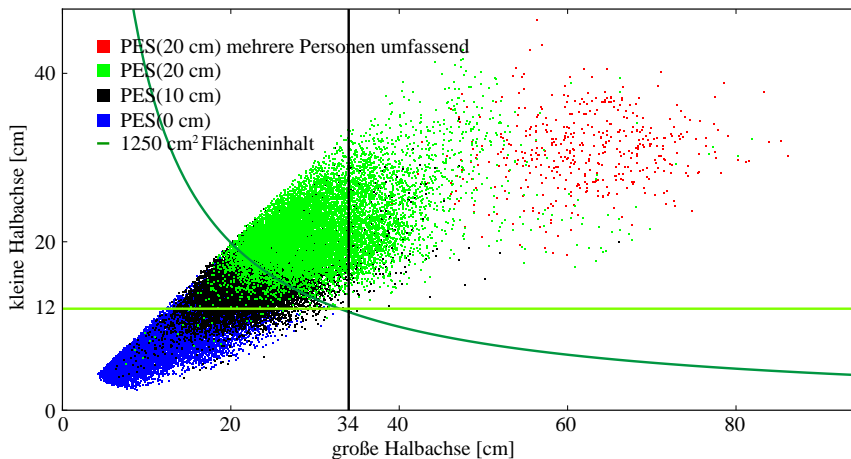


Abb. 3.59: Kombinationen von Halbachsen der Ellipsen in den PES eingefärbt gemäß Abstand zur jeweiligen Personen-größe; Linien zeigen entsprechend ihrer Farbe Grenzen der Fälle zur Identifikation eines PES als Person (siehe Seite 112); rote Punkte gehören zu Ellipsen der PES(20 cm), die mehr als eine Person umschließen, grüne Punkte zu Ellipsen der PES(20 cm), die nur einer Person zugeordnet sind

PES kleiner [cm]	falsch positiv	falsch negativ
15	0,7% (10)	2,4% (34)
20	0,7% (9)	2,9% (41)
25	0,5% (7)	4,3% (62)

Tab. 3.3: Erkennungsfehler in Abhängigkeit von einer geforderten Mindestgröße der PES; Falsch-Ermittlungs- und Falsch-Negativ-Rate von 1377 Personen in einem dichten (ca. 1,7 Personen/m² ähnlich Abb. 4.18) Personenstrom (reiner Erkennungsschritt ohne Tracking), vgl. Tab. 3.7

Die Minimalhöhe von 20 cm, wie auch Punkt 4, erlaubt PES, die nur den Kopf beschreiben. Die Schulterhöhe beginnt im Schnitt 30 cm unterhalb der Körpergröße (siehe Abb. 3.3 in [492]).

Als Annahme für die Position einer Person wäre bei optimalen Höhenlinien der Mittelpunkt der höchsten Ellipse des jeweiligen PES denkbar, die im Schnitt eine halbe Schrittweite unterhalb der Personengröße liegt. Aber gerade die erste oberste Höhenlinie respektive Ellipse kann an sehr unterschiedlichen Positionen auf dem Kopf liegen, da sie je nach Personengröße die erste kleine detektierte Erhöhung aufzeigt und somit u. a. stark von der Kopfneigung abhängt. Zudem beeinflusst die Ungenauigkeit im Disparitätsfeld die Lage stark (siehe Abb. 3.50 d und 3.51 d). Daher wird der Mittelpunkt von PES(10 cm) (ca. Augenhöhe) gewählt (siehe Abb. 3.50 e und 3.51 e). Diese Ellipse hat innerhalb des PES die geringste Standardabweichung in ihren Radien (durchschnittlich über alle Blickwinkel $\pm 4,5$ cm bzw. $\pm 2,9$ cm, siehe Abb. 3.57) und unter jedem Blickwinkel eine ähnliche Größe (Steigung der Regressionsgeraden in Abb. 3.57 von $0,04 \text{ cm}/^\circ$ bzw. $0,02 \text{ cm}/^\circ$), was einer stabilen mit wenig Schwankungen versehenen Trajektorie zu Gute kommt. Die gewählte Ellipse entspricht in Abb. 3.50 d bzw. 3.51 d der rot gefärbten Ellipse, die die zweite im PES mit einer Schrittweite von 10 cm ist.

Abb. 3.60 zeigt eine Seitenansicht des durchschnittlichen PES aller 402 Pyramiden, die bei einem Hermes-Versuch (aus Versuchsreihe Nr. 6) in einem maximalen Winkel zur optischen Achse von $5,7^\circ$ erzeugt wurden. Als Schrittweite der Höhenlinien ist 5 cm gewählt worden. Die hellblauen Linien zeigen den Durchmesser entlang der großen Halbachse bzw. Hauptachse, die dunkelblauen Linien den Durchmesser entlang der klei-

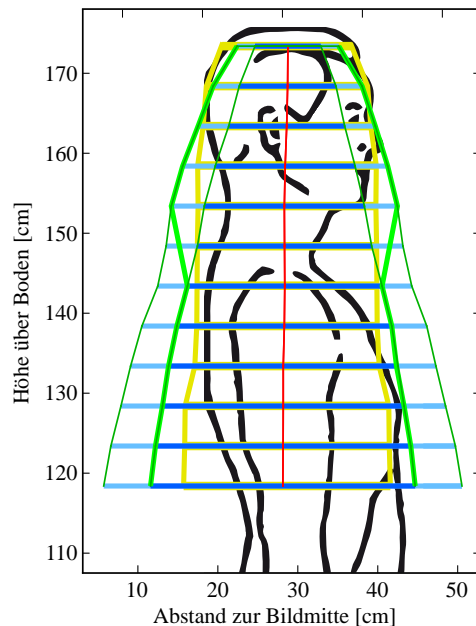


Abb. 3.60: Seitenansicht des durchschnittlichen PES eines Versuches aus Versuchsreihe Nr. 6 unter Berücksichtigung aller PES innerhalb eines maximalen Winkels zur optischen Achse von $5,7^\circ$; die Maximalhöhe aller PES wurde auf $5/2$ cm unterhalb der Durchschnittsgröße aller Personen normiert; hellblaue Linien entsprechen der Hauptachsen, dunkelblaue Linien der Nebachsen; die rote Linie verbindet die Mittelpunkte der Ellipsen, die dunkelgrünen Linien laufen entlang der Hauptachsen, wobei die dicken die für die abgebildete Seitenansicht der Person relevanten Achsen hervorhebt; hinterlegt ist die PES mit einer Durchschnittsperson gemäß [493]; die gelbe Linie zeigt den Verlauf des optimalen PES der abgebildeten Person

nen Halbachse bzw. Nebenachse. Die rote Linie verbindet die Mittelpunkte der Ellipsen, die dünnen dunkelgrünen Linien laufen entlang der kleinen und großen Hauptachsen und zeigen damit das Höhenprofil. Jeder PES ist vor der Durchschnittsbildung dahingehend normiert worden, dass die oberste Ellipse auf die Höhe der durchschnittlich ersten Schnittebene bei 176 cm - 5/2 cm verschoben wurde. Hinterlegt wurde die Grafik mit einer Person mit den durchschnittlichen Maßen eines erwachsenen Deutschen [493]. Die dicke grüne Linie zeigt an, welche Halbachsen für die abgebildete Person relevant sind, da die Lage der Hauptachse sich entlang der Körperachse in Halshöhe um 90° dreht. Das Profil der Hauptachse weist in unterer Halshöhe eine Abschwächung der Ausweitung auf. Der gelbe PES zeigt den Verlauf für die Pyramide, wie sie für die abgebildete Person optimal aussehen würde. Sie weist kaum eine Aufweitung nach unten hin auf, da die Sehstrahlen des Stereosystems unter dem hier betrachteten Winkel nahezu orthogonal zur Bewegungsebene verlaufen.

Durchschnittliche PES entsprechend Abb. 3.60 sind gruppiert nach Betrachtungswinkel (*binning*) in Abb. 3.61 abgebildet. Hierbei ist neben der Normierung auf Personengröße zudem jeder PES horizontal verschoben, so dass der Mittelpunkt von PES(10 cm) in der Mitte des Winkelintervalls liegt. Die Intervalle teilen das gesamte Spektrum der Betrachtungswinkel in sechs äquidistante Teilintervalle auf. Die metrische x-Achse zeigt den Abstand zur Bildmitte, um die Ausdehnung der Halbachsen ablesen zu können. Abb. 3.61 zeigt die relative Lage der Mittelpunkte der PES zueinander, die bei zunehmendem Betrachtungswinkel weiter auseinander liegen. Zudem weiten sich die PES bei steigendem Winkel stärker auf, was auch den Abb. 3.63 und 3.57 zu entnehmen ist. Die gelben Geraden hinter den roten Verbindungslinien der Mittelpunkte der durchschnittlichen Ellipsen ist eine lineare Regression der Mittelpunkte. Die Regressionsgeraden korrelieren sehr gut mit den Mittelpunkten und können als Mittelachsen der PES aufgefasst werden (Korrelationskoeffizient nach Pearson für alle Mittelachsen größer als 0,999; zur Korrelationsbestimmung wurden die Datenpunkte aufgrund des nahezu senkrechten Verlaufs um 45° gedreht). Die Steigung der Geraden wird bei ansteigendem Winkel flacher.

Abb. 3.62 zeigt den Zusammenhang zwischen Blickwinkel und der Steigung der Mittelachsen der durchschnittlichen PES. Hierfür wurde der gesamte Wertebereich der vorkommenden Blickwinkel α in 20 äquidistante Intervalle aufgeteilt und der Kehrwert der Steigung der Mittelachsen der durchschnittlichen PES dieser Intervalle als blaue Punkte aufgetragen. Die rote Regressionsgerade

$$-0,04 + 0,019\alpha/^{\circ}$$

korreliert gut (Korrelationskoeffizient nach Pearson von 0,996) mit den Wertepaaren und könnte genutzt werden, um

- aufgrund der ermittelten Achse in Abhängigkeit vom Winkel eine Person zu identifizieren oder als falsch positiven PES abzuweisen,
- die Position einer Person stabiler und genauer als über den Mittelpunkt von PES(10 cm) festzulegen; denkbar wäre die Ermittlung der Mittelachse jedes PES, bei der bei der Regression die Mittelpunkte jeder Ebene nach der durchschnittlichen Standardabweichung der Radian in der Ebene (siehe Abb. 3.57) oder der durchschnittlichen Standardabweichung des Mittelpunktes von der durchschnittlichen Mittelachse gewichtet werden könnten, und die Nutzung des Schnittpunktes der ermittelten Mittelachse mit PES(0 cm) als Position für die jeweilige Person oder
- die aufgrund der perspektivischen Sicht zur Bildmitte hin fälschlich verschobene Position einer Person von der optischen Achse wegzubewegen (vgl. Abschnitt 3.6.6).

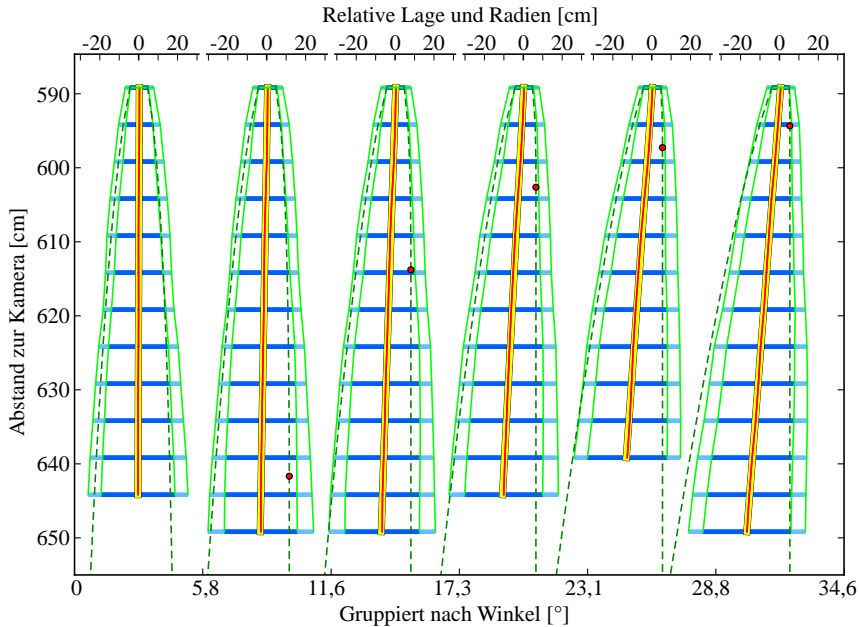


Abb. 3.61: Seitenansicht durchschnittlicher PES zusammengefasst in sechs äquidistante Intervalle von Blickwinkeln, unter denen der Mittelpunkt von PES(10 cm) zu sehen ist; die abgebildeten Elemente entsprechen denen in Abb. 3.60; Normierung aller Pyramiden auf eine Durchschnittsentfernung und einer horizontalen Verschiebung, um die Mittelpunkte von PES(10 cm) in das Zentrum des jeweiligen Winkelintervalls zu bewegen; die relative Lage der Mittelpunkte verschiebt sich bei steigendem Winkel, so dass die Steigung der gelben Regressionsgeraden flacher wird (siehe Abb. 3.62); die PES sind für größere Winkel nach unten hin breiter; die dunkelgrünen gestrichelten Linien zeigen die Lage eines elliptischen Rotationsparaboloids unter entsprechendem Blickwinkel; ab dem roten Punkt ist die Rückseite des Paraboloids nicht mehr erkennbar und stagniert somit die optische Ausdehnung

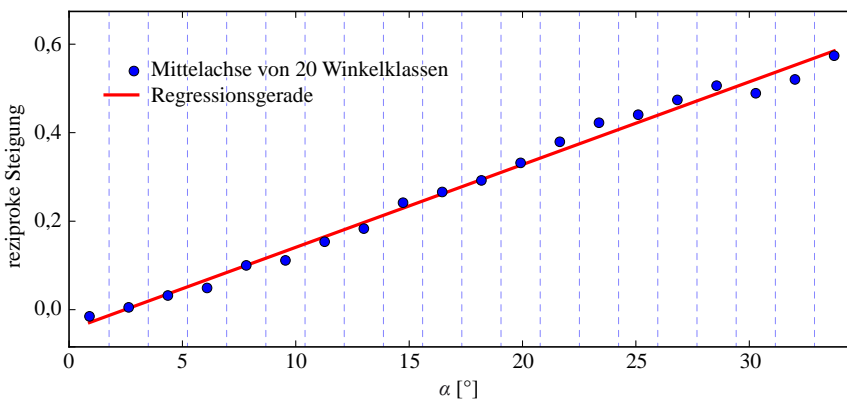


Abb. 3.62: Reziproke Steigung der Mittelachsen von durchschnittlichen PES in 20 Winkelklassen von $1,68^\circ$ (vgl. gelbe Regressionsgeraden in Abb. 3.61); rote Regressionsgerade bei $-0,04 + 0,019\alpha/^\circ$

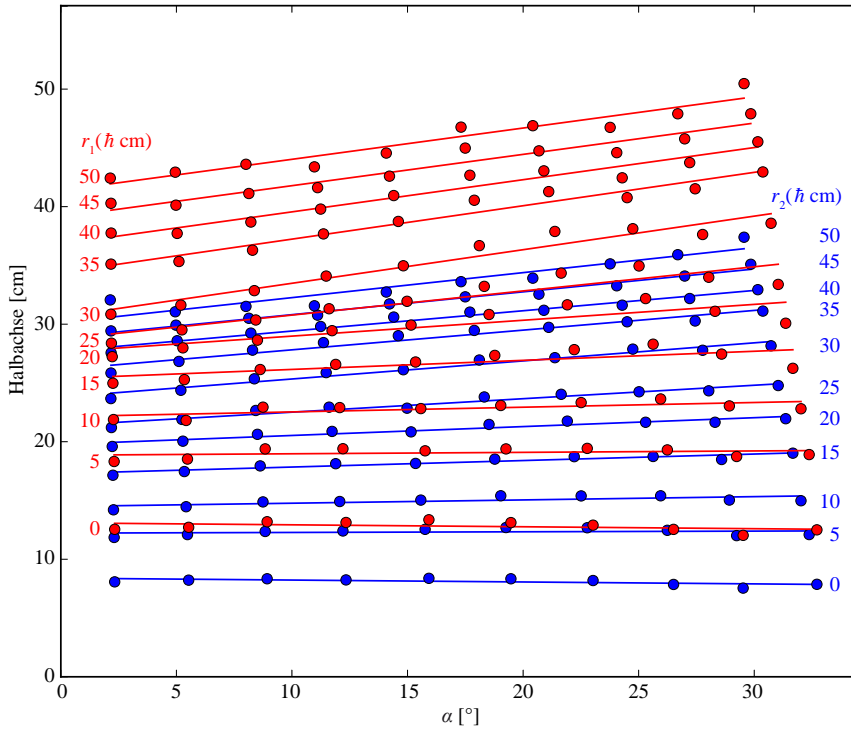


Abb. 3.63: Durchschnitt der Länge der großen (rot) und kleinen (blau) Halbachse gruppiert je Höhe nach 10 gleich großen Intervallen von Blickwinkeln bzw. Winkel zur optischen Achse von α für die 11 obersten Höhenlinien im Abstand von 5 cm jeder Person; die Regressionsgeraden zeigen die wachsenden Längen der Halbachsen für steigende Blickwinkel; dieser Anstieg ist für tieferliegende Höhenlinien approximierende Ellipsen innerhalb der PES größer; Abb. 3.57 zeigt für eine Auswahl die Standardabweichung

Umgesetzt wurde dies in `PeTrack` noch nicht und findet somit keine Anwendung bei den Ergebnissen im folgenden Abschnitt.

Die dunkelgrünen gestrichelten Linien in Abb. 3.61 zeigen die Pixelposition des Querschnitts eines elliptischen Rotationsparaboloids mit einem Leitkoeffizienten von -0,12 und einem Achsenabschnitt in durchschnittlicher Kopfhöhe für den mittleren Blickwinkel des entsprechenden Winkelintervalls. Ab dem roten Punkt ist die Rückseite des Paraboloids nicht mehr erkennbar, so dass die optische Ausdehnung aufgrund der perspektivischen Sicht stagniert. Die Achse dieses einfachen Körpers fällt bei steigendem Winkel schneller ab und zeigt, dass dieser sich nicht an die real gemessene Form des Körpers anpassen lässt. Dies gilt auch für weitere einfache Formen wie Quader unterschiedlicher Dicke. Das komplexe Zusammenspiel von Körpergeometrie, geglättetem Höhenfeld und den die Höhenlinien approximierenden Ellipsen (siehe Abb. 3.81 c und d) ist nicht durch diese Formen reproduzierbar, so dass diese zur Identifikation nicht herangezogen werden können.

Für die Abb. 3.63 und 3.57 wurde das Gesamtintervall der Betrachtungswinkel je Ebene im PES in zehn äquidistante Bereiche aufgeteilt, in denen sich außer im ersten und letzten Teilintervall in jeder Ebene mehr als 1000 Ellipsen befinden (minimal 390). Insgesamt handelt es sich um 12699 Pyra-

miden mit 121378 Ellipsen aus 617 Einzelbildern eines Versuchs aus Versuchsreihe Nr. 6, bei dem alle Personen die Hälfte der horizontalen und die Hälfte des vertikalen Bildes durchlaufen mussten (entsprechend Abb. 4.28). Ellipsen, die mehr als eine Person umschließen und somit mehreren PES zugeordnet sind, wurden nicht berücksichtigt. Da das Intervall je Ebene unterteilt wurde, wandert der durchschnittliche Winkel für tieferliegende Ebenen innerhalb der PES zur Bildmitte bzw. an die optische Achse, da der Winkel, unter dem diese Ellipsen zu sehen sind, kleiner ist. Die Länge der großen (rot) und kleinen (blau) Halbachse wurde ebenenweise durch eine Regressionsgerade approximiert. Höherliegende Ebenen im PES besitzen bei der Approximation kleinere Fehler und haben kaum steigende Längen der Halbachsen. Zwischen PES(15 cm) und PES(30 cm) (unterer Kopfbereich bis zu den Schultern) nimmt die Steigung der Längen bei steigendem Winkel zu, nach PES(30 cm) (unterhalb Schulterhöhe) bleibt die Steigung ungefähr bei $0,27\text{cm}/^\circ$ bzw. $0,19\text{cm}/^\circ$. Die Streuung um die Regressionsgerade ist entlang der Teilintervalle nicht für alle Ebenen des PES identisch. Die geringste Standardabweichung besitzen die Ellipsen von PES(10 cm) für beide Halbachsen.

Abschließend lässt sich sagen, dass die aufgeführten heuristischen Werte sich an die Körperproportion erwachsener Deutscher anlehnen (siehe [296]) und zum Teil vom geglätteten perspektivischen Höhenfeld beeinflusst werden. Dabei wurden die Werte so gewählt, dass, wie bereits zuvor, streng ausselektiert wird, da eine Person nicht in jedem Frame erkannt werden muss, sondern zwischenzeitlich nur verfolgt werden kann.

Zur weiteren Verringerung der falsch positiven Detektionen können die in Abschnitt 3.3.5 vorgestellten Verfahren während des Trackings herangezogen werden. Gesamtergebnisse sind in Abschnitt 3.6.6 zu finden.

3.6 Ergebnisse

Viele Ergebnisse wurden bereits in den Erläuterungen zu den Schritten der Extraktion der Laufwege genannt, wenn diese z. B. zum Verständnis der Wahl von Parametern nötig waren. Hier sollen darüber hinaus folgende Punkte genauer untersucht werden:

- Fehler durch eine ungenügende Kalibrierung,
- Genauigkeit der über die Disparität ermittelten Personengrößen,
- Qualität der mittels Strukturelementen, Farbmarker und des markerlosen Verfahrens – auch beim Einsatz während einer Feldstudie – ermittelten Laufwege,
- Verfahrens- und Parameterwahl zur Bestimmung der Disparität für das markerlose Verfahren und
- Rechenzeitbedarf der Schritte zur Ermittlung der Laufwege in Abhängigkeit vom Verfahren und untersuchten Szenario.

Dabei sollen hier nur Ergebnisse der reinen Extraktion beschrieben werden. Erkenntnisse, die aus den Laufwegen gewonnen wurden, werden exemplarisch in Abschnitt 4.4 aufgeführt.

Zur Bewertung der Ergebnisse werden je nach betrachteten Größen bzw. Verfahren unterschiedliche Maße eingeführt, die den kritischen Aspekt am ehesten wiedergeben können. Diese sind z. B. die Falsch-Ermittlungs- und Falsch-Negativ-Rate einschließlich oder ausschließlich des verbindenden

Trackings, der Abstand des ermittelten Laufweges von der optimalen Position, sowie die abgeleitete Geschwindigkeit oder Beschleunigung mit unterschiedlicher Mittelung, um systematische Fehler oder die Glattheit zu untersuchen. In [364, 495] werden Metriken zur Bewertung von Laufwegen beschrieben, wobei in diesen Arbeiten zum einen einzeln betrachtete Maße zusammengefasst werden, um für Vergleiche einen Gesamtwert vorliegen zu haben, zum anderen Probleme der Fehlzuordnung bei der Erkennung zu vorhandenen Personen bewertet werden, die für die hier betrachteten Überkopfaufnahmen und entsprechender Bildwiederholrate irrelevant sind, oder das verbindende Tracking ganz unterschlagen wird. Einzig die MOTP-Metrik (*multiple object tracking precision metric*) aus [364] entspricht dem in Gleichung 3.40 eingeführten Fehler.

Zur Beurteilung eines Klassifikators, z. B. einer Methode zur Detektion einer Person, werden relative Häufigkeiten von Fehlern als quantitatives Maß herangezogen. Sei t_p (*true positives*) die Anzahl der korrekt gefundenen Personen, f_n (*false negatives*) die Anzahl der nicht gefundenen Personen und f_p (*false positives*) die Anzahl der fälschlich gefundenen Personen. Folgende Bezeichnungen finden hier Verwendung [496, 497]:

- Richtig-Positiv-Rate (*true positive rate*) oder auch Sensitivität (*sensitivity*) oder Trefferquote (*recall, hit rate*) gibt den Anteil der korrekt als positiv klassifizierten Objekte an der Gesamtheit der tatsächlich positiven Objekte an bzw. entspricht dem Anteil der korrekt detektierten Personen an allen vorkommenden Personen:

$$t_p / (t_p + f_n),$$

- Falsch-Negativ-Rate (*false negative rate, miss rate*) ist der Anteil der falsch negativen (*false negatives*), also fälschlich als negativ klassifizierten Objekte, die in Wirklichkeit positiv sind, bzw. der Anteil der nicht erkannten Personen an allen vorkommenden Personen:

$$1 - t_p / (t_p + f_n) = f_n / (t_p + f_n),$$

- Richtig-Negativ-Rate (*true negative rate*) oder auch Spezifität (*specificity*) oder korrekte Ablehnungsrate (*correct rejection rate*) gibt den Anteil der korrekt als negativ klassifizierten Objekte an der Gesamtheit der in Wirklichkeit negativen Objekte, bzw. der korrekt als Person abgelehnten fälschlich vorgeschlagenen Vorkommnissen einer Person an,
- Falsch-Positiv-Rate (*false positive rate*) oder auch Ausfallquote (*fallout*) ist der Anteil der falsch positiven (*false positives*), also fälschlich als positiv klassifizierten Objekte, die in Wirklichkeit negativ sind, an der Gesamtheit der in Wirklichkeit negativen Objekte, bzw. der Anteil fehlerhaft erkannter Personen an allen fälschlich vorgeschlagenen Vorkommnissen einer Person,
- Positiver Vorhersagewert (*positive predictive value*) oder auch Relevanz, Wirksamkeit oder Genauigkeit (*precision*) gibt den Anteil der korrekt als positiv erkannten Objekte an der Gesamtheit der als positiv erkannten Objekte bzw. den Anteil der richtig erkannten Personen an allen durch das Verfahren ermittelten Vorkommnissen an:

$$t_p / (t_p + f_p),$$

- Falsch-Ermittlungs-Rate (*false discovery rate*) ist der Anteil der falsch positiven (*false positives*), also fälschlich als positiv klassifizierten Objekte, die in Wirklichkeit negativ sind, an

der Gesamtheit der tatsächlich als positiv klassifizierten Objekte, bzw. der Anteil der fehlerhaft erkannten Personen an allen durch das Verfahren ermittelten Vorkommnissen:

$$1 - t_p / (t_p + f_p) = f_p / (t_p + f_p).$$

Die Richtig-Negativ- und Falsch-Positiv-Rate können im Allgemeinen nicht angegeben werden, da in jedem Verfahrensschritt die Grundgesamtheit der möglichen Vorkommen einer Person variiert. Diese Raten wären nur für einzelne Schritte wie der Identifikation mittels der PES in Abschnitt 3.5.5 angebbbar und könnten in diesem Fall die Quote der richtig oder falsch abgelehnten PES quantifizieren. Stattdessen wird für die Bewertung eines Verfahrens die Falsch-Ermittlungs-Rate verwandt, die die falsch positiven Detektionen im Verhältnis zu den durch das Verfahren ermittelten Personen setzt.

In einigen Arbeiten werden auch die von statistischen Tests gebräuchlichen Fehler 1. Art (*Type I error*), der fälschlichen Annahme der Alternativhypothese, und Fehler 2. Art (*Type II error*), der fälschlichen Ablehnung der Alternativhypothese, zur Bezeichnung der Falsch-Positiv- und Falsch-Negativ-Rate verwandt.

3.6.1 Kalibrierung

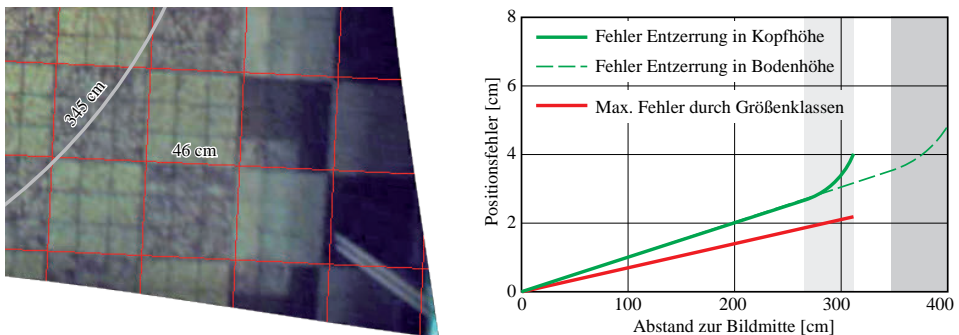


Abb. 3.64: Positionsfehler bei Kasernen-Versuchen (Nr. 3); (links) Bildausschnitt rechts unten von Abb. 3.2 c, bei dem die Abweichung des mittels PeTrack überlagerten roten Gitters vom 46 cm-Raster der Bodenfliesen insbesondere zum Rand und zu den Ecken zu beobachten ist; grauer Kreisausschnitt im Abstand von 345 cm (Bodenhöhe) zur Bildmitte; (rechts) bis 275 cm (Kopfhöhe) bzw. 345 cm (Bodenhöhe) ungefähr lineare Zunahme des durch die fehlerbehaftete Entzerrung verursachten Fehlers in der Positionsbestimmung; im grau hinterlegten Bereich zum Rand des Bildes steigt dieser Fehler überproportional an; die rote Linie beschreibt den in Abhängigkeit vom Abstand zur Bildmitte linear ansteigenden maximalen Fehler durch die Nutzung von 5 cm-Größenklassen

Neben den Fehlern bei der Erkennung einer Person sind die Ungenauigkeiten in der Positionsbestimmung eine weitere Fehlerquelle. Der Einfluss der Art des Markers dabei wird in den Abschnitten 3.6.3 ff. beschrieben. In diesem Abschnitt soll die Abhängigkeit der Genauigkeit in der Positionsbestimmung von der Kalibrierung besprochen werden.

Die Kalibrierung dient zum einen der Entzerrung (siehe Abschnitt 3.1.2) und zum anderen der Vermessung (siehe Abschnitt 3.1.1) zur Positionsbestimmung (siehe Abschnitt 3.4).

In dieser Arbeit und der Implementation in PeTrack wird bei der Entzerrung das Modell der Lochkamera um radiale und tangentielle Verzerrungsparameter bis zur Ordnung vier gemäß [341, 459] erweitert. Dabei werden die entzerrten Pixelkoordinaten entsprechend Gleichung 3.7 über vier Parameter

bestimmt. Diese Abbildung bzw. das erweiterte Modell erfasst zwar wichtige Verzeichnungen, kann aber nicht alle Aberrationen eines realen Linsensystems erfassen. Dabei unterscheidet sich die Güte der Kalibrierung je nach Optik und der Möglichkeit, diese durch das Modell abbilden zu können. Scherungen und höherdimensionale radiale Verzerrungen wurden nicht berücksichtigt. Der dadurch verursachte Fehler steigt typischerweise überproportional bei Verringerung der Brennweite. Die weitwinkligen Aufnahmen sind jedoch oft nötig, um den zu erfassenden Bereich abzudecken (siehe Abschnitt 4.1.1.2).

Der Fehler aufgrund ungenügender Entzerrung steigt von der Bildmitte zum Rand an. Für die Kasernen-Versuche (Nr. 3) ergibt sich ein Fehlerverlauf der ermittelten Position einer Person in der Bewegungsebene gemäß Abb. 3.64. Die durchgezogene grüne Linie zeigt die für die Ermittlung der Trajektorien relevante Steigung des Fehlerverlaufs von der Bildmitte bis zum Rand (Beginn des hellgrauen Bereichs bei 275 cm) und bis in die Ecken (Ende des hellgrauen Bereichs bei 310 cm) in durchschnittlicher Kopfhöhe der Personen. Die gestrichelte grüne Linie zeigt den Positionsfehler auf dem Boden, der auch im linken Bildausschnitt (rechte untere Ecke der Abb. 3.2 c) zu beobachten ist. Der Beginn des dunkelgrauen Bereichs bei 345 cm im Diagramm ist im Bild mit einem Kreisausschnitt gekennzeichnet. Das überlagerte rote Gitter im Raster von 46 cm weicht zur Ecke immer stärker vom entsprechenden Raster der Bodenfliesen ab.

Der nahezu lineare Anstieg, beginnend bei einem verschwindenden Fehler in der Bildmitte senkrecht unter der Kamera bis zum grauen Bereich, hat ungefähr eine Steigung von 1 cm/m und wächst danach überproportional bis auf ca. 4 cm in den äußersten Ecken an. Damit ist dieser Fehler fast 50% größer als der Maximalfehler, der durch die 5 cm-Größenklassen, die bei den Kasernen-Versuchen den Fehler durch die perspektivische Verzerrung begrenzen, verursacht wird (rote Gerade in Abb. 3.64, siehe auch Abschnitt 3.4.1). Dieser in Gleichung 3.20 eingeführte Fehler e_h monokularer Kameras kann ebenfalls in Abhängigkeit von der Entfernung $\|p_i(t) - \vec{c}'\|$ zur Bildmitte \vec{c}' und der senkrechten Distanz zur Kamera $a - \bar{p}^h$ angegeben werden:

$$e_h = \frac{d_h \|p_i(t) - \vec{c}'\|}{2(a - \bar{p}^h)} = \frac{5}{2(535 - 180)} \approx 0,7 \text{ cm/m}.$$

Der Gesamtfehler aufgrund der Größenklassen und der fehlerbehafteten Entzerrung von ca. 1,7 cm/m, mit einem überproportionalen Anstieg zum Rand mit einem Maximalfehler von 4,7 cm in den äußersten Ecken, ist dahingehend zu relativieren, dass für die interessierende Position in der Bewegungsebene, u. a. durch das unterschiedliche Tragen der Kopfbedeckung oder Bewegungen des Kopfes unabhängig zur Hauptbewegungsrichtung, bereits Fehler ähnlicher Größenordnung verursacht werden.

Dabei haben die beiden Fehler unterschiedliche Qualitäten. Steigt bei der Zusammenfassung von Personengrößen zu Größenklassen im Gegensatz zur exakten Bestimmung der Personengröße nur die Ungenauigkeit in der Positionsbestimmung, so stellt die fehlerbehaftete Entzerrung einen systematischen Fehler dar, der sich auch anders auf die für die Fußgängerdynamik relevanten Größen wie u. a. Personendichte und Geschwindigkeit auswirkt (siehe Abschnitt 3.6.9).

Neben dem Fehler, der aus einem eingeschränkten Modell zur Entzerrung resultiert, kann es auch bei der Bestimmung der Verzerrungsparameter selbst gemäß Gleichung 3.8 zu Fehlern kommen, wenn z. B. die angenommene Lage des verwendeten Musters aufgrund von leichten Unebenheiten in der konkreten Repräsentation, z. B. einem auf eine Platte aufgetragenen Ausdruck des Kalibrierungsmusters, nicht mit der realen Lage übereinstimmt. Da das Muster für gewöhnlich nah der Kamera aufgenommen wird, führen bereits kleine Unebenheiten zu merklichen Fehlern. Zur Minimierung dieses Fehlers wurde das Kalibrierungsmuster zwischen zwei Glasscheiben fixiert. Wichtig ist zudem, dass

die Bestimmung der Verzerrungsparameter oder zumindest die für die Kalibrierung benötigten Aufnahmen mit dem unveränderten Kamerasystem durchgeführt wird, da schon das leicht unterschiedliche Anschrauben der Objektive die Parameter verändern kann. Am besten werden die Aufnahmen direkt am Ort der Experimente durchgeführt, so dass u. a. auch Temperaturschwankungen keinen Einfluss auf die Ergebnisse haben.

Bei der Vermessung wird von einer senkrechten Ausrichtung der Kamera und einer korrekt bestimmten Aufhängehöhe der Kamera ausgegangen. Bei der Installation der Kameras während der Versuche wurde genau darauf geachtet, dass das Kameralot exakt auf die Bildmitte abgebildet wird. Die Aufhängehöhe muss nicht ausgemessen werden, sondern kann auch aus dem bei der Bestimmung der intrinsischen Parameter berechneten Skalierungsfaktor und der Brennweite nach Spezifizierung eines Koordinatensystems auf dem Boden bestimmt werden. Die Ausrichtung und die Skalierung des Koordinatensystems kann in *PeTrack* manuell sehr genau auf Subpixel-Ebene durchgeführt werden. Für Stereokameras ließe sich direkt aus der Disparität die 3D-Position ermitteln, so dass man nicht auf eine exakte Ausrichtung angewiesen ist. Dies wurde insbesondere in 3D-Geometrien wie Treppen genutzt und erlaubte auch in der Feldstudie mit einer leicht schrägen Aufhängung eine korrekte Bestimmung von Trajektorien. Manuelle Überprüfungen für ausgemessene Punkte im Bild ergaben zu vernachlässigende Fehler.

Die für monokulare Kameras beschriebenen Fehler bei der Kalibrierung gelten auch für die einzelnen Kameras eines Stereosystems, insbesondere ist die genaue Lage der einzelnen Kameras im Weltkoordinatensystem und damit die Lage der Kameras zueinander mit entscheidend für die Qualität der ermittelten Disparität zur Entfernungsbestimmung und insgesamt zur Bestimmung der 3D-Position eines Pixel. Die Auswirkung dieses Fehlers wird am Beispiel der Größenbestimmung der Versuchspersonen im folgenden Abschnitt beschrieben.

3.6.2 Größenbestimmung mittels Disparität

Der Fehler, der bei Vernachlässigung der individuellen Personengröße gemacht wird (z. B. Farbmarker Nr. 4 ohne Nutzung von Stereoaufnahmen), wird in Abschnitt 3.4 beschrieben. Dieser Fehler wurde z. B. bei den Japan-Versuchen (Nr. 8) in Kauf genommen, dort aber durch eine sehr hohe Kameraaufhängung von 2033 cm – 2044 cm verringert, da damit durch einen möglichen Zoom (Vergrößerung der Brennweite) bzw. gleichbedeutend mit der Nutzung nur eines Bildausschnittes der für den Fehler verantwortliche Winkel zum Kameralot reduziert werden kann.

Die Steigerung der Genauigkeit bei der Positionsbestimmung einer Person, die mit einer Zuordnung aller Personen zu Größenintervallen einhergeht, kann Abschnitt 3.4.1 entnommen werden. In diesem Abschnitt soll genauer betrachtet werden, wie die Güte der Größenbestimmung über die Disparität ist.

Abschnitt 3.4.3 zeigt die möglichen Fehler bei der Bestimmung der Disparität auf. Gleichung 3.26 gibt die Abhängigkeit des Fehlers in der Entfernungsbestimmung vom Fehler in der Disparität an. Der Korrespondenzfehler für beide bei den Hermes-Versuchen (Nr. 6) eingesetzten Stereokameras kann damit gemäß Abb. 3.65 aufgrund des dort abzulesenden systematischen Fehlers in der Entfernungsbestimmung von ca. $\Delta z'_0 = 10\text{cm}$ wie folgt angegeben werden:

$$\Delta\delta \approx \frac{\Delta z'_0 b f'}{z_0'^2} = \frac{10\text{cm} \cdot 983\text{px} \cdot 24\text{cm}}{(608\text{cm})^2} \approx 0,64\text{px}. \quad (3.28)$$

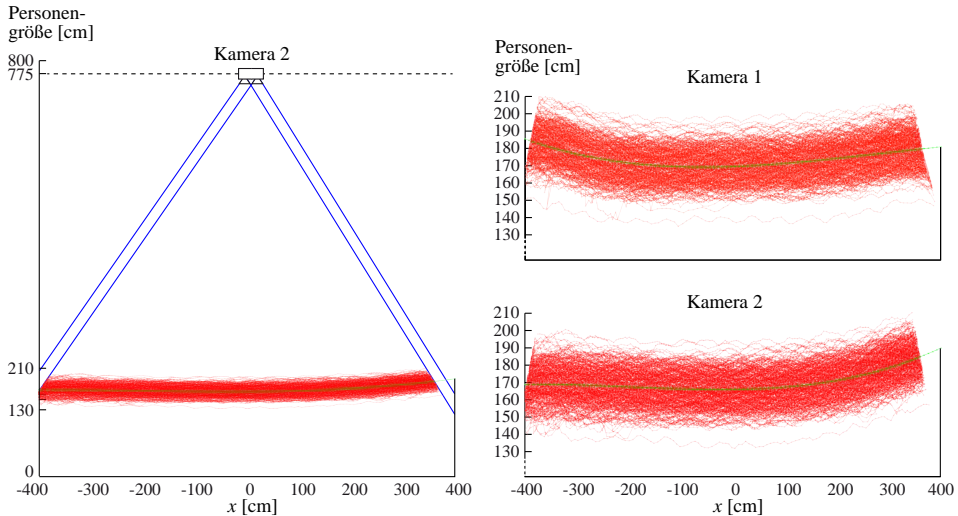


Abb. 3.65: Personengröße entlang der Hauptbewegungsrichtung eines ebenen Versuches aus Versuchsreihe Nr. 6, ermittelt durch zwei Stereokameras vom Typ Nr. 4; (links) gleich skalierte Achsen; Position sowie Basislänge des Stereokamerasystems sind maßgetreu eingezeichnet; (rechts) überhöht dargestellte Personengrößen beider Stereokamerasysteme zeigen den ungleichmäßigen Verlauf der Größe; die grüne Linie interpoliert die mittlere Größe entlang der Hauptbewegungsrichtung durch ein kubisches Polynom

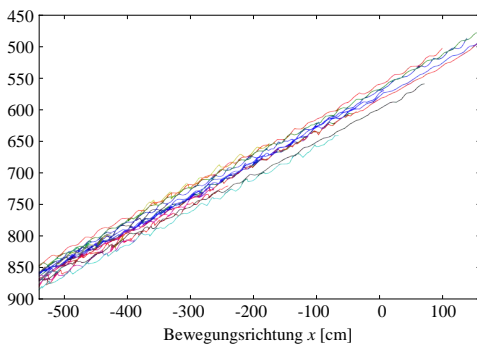


Abb. 3.66: Laufwege von auf einer Treppe herunterlaufenden Personen aus Versuchsreihe Nr. 6

Die in Abb. 3.65 zu sehenden Fehler resultieren am wahrscheinlichsten aus den Punkten Nr. 7 und 8 (ungenauere Entzerrung und Rektifizierung) der Liste der Probleme bei der Disparitätsbestimmung von Seite 96 und nicht von einer fehlerhaften Korrespondenzanalyse, da, wie in Abb. 3.68 a zu sehen, sogar die aufgrund der Schritte schwankende Personengröße detektiert wird und auch der Kopfverlauf beim Treppenabstieg die Stufen erkennen lässt (siehe Abb. 3.66). Zudem lassen die matten kontrastreichen Marker eine Fehlzuordnung der Bildpunkte beider Kameras unwahrscheinlich erscheinen. Des Weiteren steigt der Fehler systematisch und zum Rand hin an und ist für alle Verfahren der Liste auf

Seite 91 gleich.

Die Werte der intrinsischen und die Lage der einzelnen Kameras im Stereokamerasystem zueinander betreffenden extrinsischen Parameter werden jedoch bei den eingesetzten Stereokameras vom Hersteller im Werk bestimmt und sind fest in das Stereokamerasystem eingeschrieben. Der Kamerahersteller gibt den maximalen Fehler in der Disparität seiner Kameras mit $\Delta\delta = 0,2 \text{ px}$ an [498, 499], aber trotz eines Einschickens und Neuvermessens wurde der Fehler nicht geringer und liegt somit drei Mal so hoch wie vom Hersteller angegeben. Der Hersteller hat das Problem bestätigt und führt es auf Schwankungen der Temperatur während des Transports zurück.

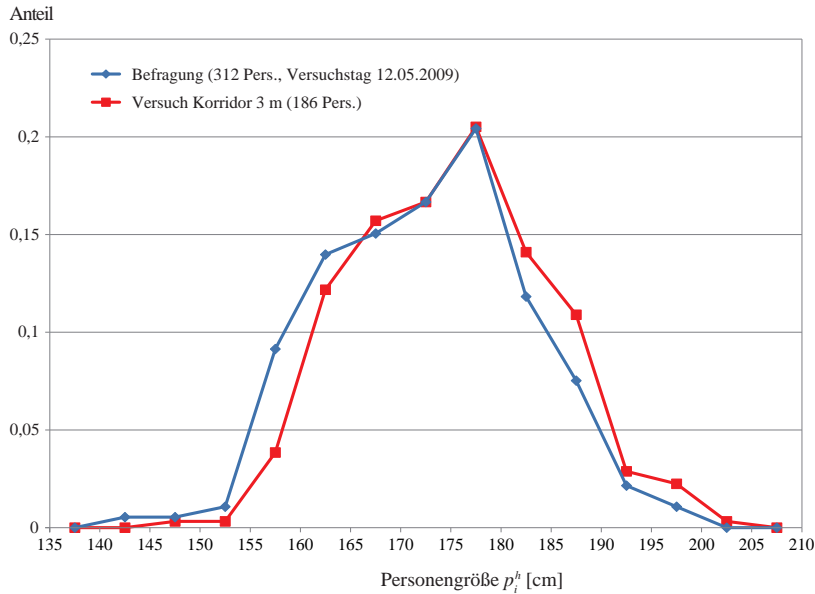


Abb. 3.67: Verteilung der Personengröße p_i^h laut Befragung aller an den an diesem Tag durchgeführten Hermes-Versuchen (Nr. 6) beteiligten Personen (blau) und die mittels der Stereokamera automatisch über die Disparität erfassten Größen bei einem dieser Versuche (rot)

Um diesem Problem zu begegnen, besteht in *PeTrack* die Möglichkeit neben den aus der Disparität bestimmten Personengrößen auch einen Mittelwert pro Person herauszuschreiben. Dies ist wohlge-merkt nur eine Hilfe bei ebenen Versuchen und könnte einen systematischen Fehler in den Laufwegen z. B. auf einer Treppe nicht eliminieren. Darüber hinaus kann in *PeTrack* die Personengröße beim Exportieren neu berechnet, eingeschränkt, geglättet oder ganz verworfen werden (siehe Abb. D.5).

Dass trotzdem die über die Disparität ermittelten durchschnittlichen Personengrößen gut mit der Rea-lität übereinstimmen ist Abb. 3.67 zu entnehmen. Die tatsächliche durchschnittliche Körpergröße von $\bar{p}^h = 176\text{cm} \pm 9\text{cm}$ wurde durch eine Befragung ermittelt. Das Hinzuaddieren eines Offsets zur automatisch ermittelten Personengröße würde die Deckung von Befragung und Messung noch wei-ter erhöhen. Es ist aber zu beachten, dass die Versuchspersonen oft nicht genau ihre aktuelle auch über den Tag hinweg schwankende Größe wissen (können) und zudem diese für gewöhnlich ohne das Tragen von Schuhen ermittelt haben.

Die bereits erwähnte gute lokal bestimmte Disparität kann der Auf- und Abbewegung der Personen während des Gehens in Abb. 3.65 oder Abb. 3.68, aber auch der Nachempfindung des Treppenstei-gens in Abb. 3.66 entnommen werden. Dabei korreliert die Schwankbewegung der Person beim Ver-lagern des Körpergewichtes von einem auf das andere Bein (siehe Abb. 3.21 oder Video Nr. 23) mit der Größenänderung. Im Wendepunkt der oszillierenden Bewegungskurve findet die Verlagerung des Körpergewichtes von einem auf den anderen Fuß statt, bei der beim Gehen kurzzeitig beide Beine den Boden berühren und dabei den tiefsten Punkt in einer Schrittfolge beschreiben (siehe Abb. 3.68 b). In Abb. 3.68 a ist dieser in jedem Schritt wiederkehrende Punkt durch schwarze senkrechte Linien gekennzeichnet. Die Linien zeigen die Korrelation zwischen den Wendepunkten der Bewegung

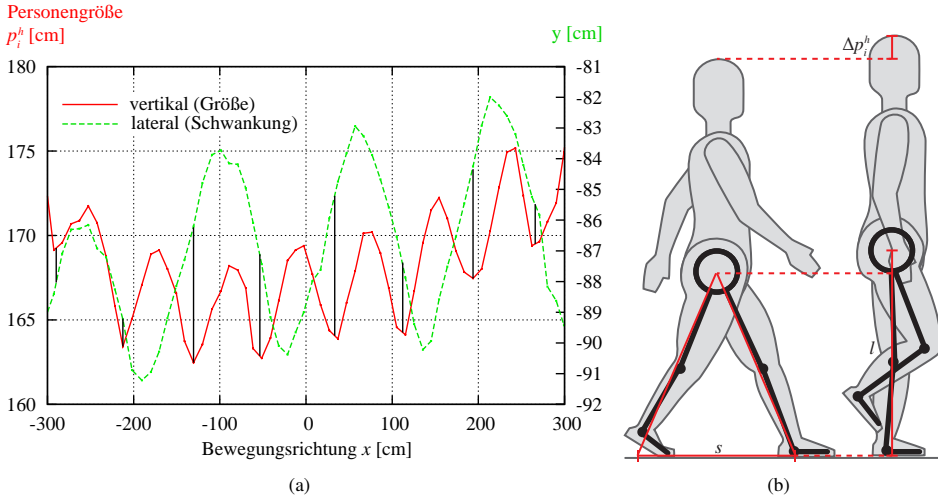


Abb. 3.68: Die Personengröße ändert sich entlang des Laufweges aufgrund des Bewegungsablaufes beim Gehen; (a) die Größe (rote Linie) korreliert mit der lateralen Pendelbewegung (grüne gestrichelte Linie) durch die wechselweise Beinbelastung; am Wendepunkt bei der Verlagerung des Körpergewichtes auf das andere Bein ist die Personengröße am geringsten (b) Stufen im Bewegungsablauf beim Gehen mit der minimalen und maximalen Personengröße; die Größendifferenz ist demnach $\Delta p_i^h = l - \sqrt{l^2 - s^2/4}$

in der x - y -Bewegungsebene und den Minima der Körpergröße während der Gewichtsverlagerung. Die Frequenz der vertikalen Größenänderung ist damit doppelt so groß wie die Frequenz der lateralen Pendelbewegung. Die Hauptbewegungsrichtung verläuft ungefähr parallel zur x - und die laterale Schwankung parallel zur y -Achse.

Abb. 3.68 b zeigt links diesen Tiefpunkt der Bewegung und rechts den Zustand der maximalen Personengröße beim Übersteigen des ausgestreckten Beines mit dem anderen Bein. Die Schwankung in der Personengröße ergibt sich damit annähernd aus der Schrittlänge s und der Beinlänge l bei Annahme eines gleichschenkligen Dreiecks der Basislänge s und Schenkellänge l zum Zeitpunkt der minimalen Personengröße mit

$$\Delta p_i^h = l - \sqrt{l^2 - \frac{s^2}{4}}. \quad (3.29)$$

Aus Abb. 3.68 a können die ungefähren Maße $s = 80$ cm, $p_i^h = 170$ cm und $\Delta p_i^h = 6$ cm einer Person in einem Experiment der Hermes-Versuche (Nr. 6) abgelesen werden. Da sich die hier benötigte projektivische Beinlänge bei den in dieser Versuchsreihe beteiligten ausgewachsenen Personen im Durchschnitt aus der Körpergröße mit $l \approx 0,45 p_i^h$ ergibt [492] müsste jedoch gelten:

$$\Delta p_i^h = 77 \text{ cm} - \sqrt{77^2 \text{ cm}^2 - \frac{80^2 \text{ cm}^2}{4}} = 11 \text{ cm}.$$

Die geringere gemessene Schwankung von ca. 6 cm könnte u. a. vom komplexeren Gangablauf über die Zehenspitzen herrühren. Die projektivische Beinlänge aller an der Versuchsreihe Nr. 6 teilgenommenen Personen in Abhängigkeit vom Alter und Geschlecht liegt laut Abb. 3.6 in [492] zwischen 79 cm und 87 cm.

Die große Schrittlänge von 80 cm für eine Person mit einer Körpergröße von 170 cm resultiert aus der für diese Person großen Gehgeschwindigkeit, da sie als eine der ersten den Versuchsaufbau durchquerte und somit ihre freie Gehgeschwindigkeit erreichen konnte. Die mit geringerer Dichte und damit höherer Geschwindigkeit steigende Schrittlänge kann [156, 500, 501] entnommen werden.

Bei der Glättung der Laufwege zur Bestimmung der Hauptbewegungsrichtung u. a. zur Berechnung der zielgerichteten individuellen Geschwindigkeit ist das Wissen über den Bewegungsablauf einer Person wichtig. Dieser Ablauf kann wie oben gezeigt neben der Pendelbewegung in der Ebene auch über die Größenänderung bestimmt werden. In seitlichen oder schrägen Aufnahmen, in denen Personen im Ganzen zu sehen sind, kann die Schrittfolge auch direkt extrahiert werden [502–504].

3.6.3 Trajektorien mit Strukturelementen markierter Personen

Für die Erkennung von Personen mit Hilfe von am Kopf befindlichen Markern mit Strukturelementen wird bei korrekter Detektion immer die gleiche Stelle des Bildes und damit der gleiche Punkt im Raum lokalisiert (siehe Abschnitt 3.2.1). Dies ist damit die genaueste Methode zur Bestimmung der Laufwege. Ein Vergleich u. a. der Positionsgenauigkeit im Verhältnis zum Farbmarker ist Abschnitt 3.6.4 und im Verhältnis zur markerlosen Detektion Abschnitt 3.6.6 zu entnehmen.

Aufgrund von möglichen hohen Dichten und gegenseitiger Verdeckung kann mit optischen Verfahren nur der Kopf lokalisiert werden. Die ermittelte Position entspricht daher nicht unbedingt dem Schwerpunkt einer Person oder projiziert auf den Boden dem Standpunkt bzw. der Position zwischen ihren Füßen. Die Trajektorie beinhaltet Kopfbewegungen, die durch die abwechselnde Beinverlagerung der Laufbewegung entsteht, aber auch durch anders motivierte Positionsänderungen, z. B. beim Umschauen oder Kopfkreisen, hervorgerufen werden. Zudem wird die relative Lage der Position des Markers durch die Körperhaltung und das Tragen der Kopfbedeckung beeinflusst. Der Kopf wird häufiger gesenkt statt der Blick nach oben gerichtet. Der geneigte Kopf könnte mit steigender Dichte häufiger vorkommen (siehe Abb. 4.1) und ist bei der Bewegung auf unebenen Untergründen (u. a. Treppen) üblich. Insgesamt ist somit von einem dichte- und vom Untergrund abhängigen Positionsfehler in Hauptbewegungsrichtung auszugehen. Dieser ist global und fällt bei der Dynamik in einer Menschenmenge nicht ins Gewicht, könnte aber die Distanz zu einer Wand in Richtung der Bewegung z. B. bei einer Engstelle beeinflussen, wenn das Senken des Kopfes vor senkrechten Hindernissen ebenfalls zur Kollisionsvermeidung durchgeführt wird (was noch zu untersuchen wäre). Für die Ermittlung der Geschwindigkeit in Hauptbewegungsrichtung ist man bestrebt, die Schwankbewegung herauszurechnen (siehe Abschnitt 4.3). Bei der Modellentwicklung hat hingegen die geschwindigkeitsabhängige Schwankung bereits erfolgreich Einfluss gefunden [55, 505], da sie den Platzbedarf einer Person maßgeblich beeinflusst.

Im Folgenden sollen einige Ergebnisse der Extraktion von mit Strukturelementen markierten Personen vorgestellt werden. Zu dieser Klasse gehören der Kasernen- (Nr. 1 gemäß Liste auf Seite 57 und Abb. 3.8 und 3.9), der Hermes- (Nr. 2), der Japan- (Nr. 3) und der Codemarker (Nr. 5). Auf den Codemarker wird dabei nicht weiter eingegangen, da die Erfassung bzw. das Auslesen nicht Gegenstand dieser Arbeit ist. Die in Tab. 3.4 aufgeführten Experimente wurden alle mit der maximalen Personenanzahl und damit größten Dichte durchgeführt.

Für ein Engstellenexperiment der Kasernen-Versuche (Versuchsreihe Nr. 3) (siehe Abb. 3.2) kam es zu einer nicht erkannten Person aufgrund von Bewegungsunschärfe. Um die Tiefenschärfe zu vergrößern und damit unterschiedlich große Personen gleich scharf abzubilden, wurde die Blende der Kamera verringert, so dass die Belichtungszeit erhöht werden musste. Dies hatte zur Folge, dass schnelle

Versuch	V. reihe ¹	Kamera ²	Marker ³	Abb.	erkannte Personen	falsch positiv	falsch negativ	Beschleunigung [m/s ²]
Engstelle	3	1	1	3.2	177	0% (0)	0,6% (1)	0,7 ± 0,7
Korridor	3	1	1	4.2 a	45	0% (0)	0% (0)	0,5 ± 0,4
Korridor	6	4	2	3.36	79	3,8% (3)	0% (0)	0,7 ± 0,6
Tribüne	6	4	2	3.3	263	1,1% (3)	0% (0)	2,2 ± 2,7
Raum	8	-	3	3.8 3	25	0% (0)	0% (0)	0,6 ± 0,5

Tab. 3.4: Erkannte Personen, Falsch-Ermittlungs- und Falsch-Negativ-Rate bei Nutzung von Markern mit Strukturelementen zur Detektionen der Versuchsteilnehmer; ¹Versuchsreihe gemäß Tab. 4.1, ²Kameramodell gemäß Tab. 4.2, ³Marker gemäß Abb. 3.8 und 3.9; die Spalte „Abbildung“ verweist auf ein Bild des zugrundeliegenden Versuchs; die Beschleunigung wurde gemäß Gleichung 3.44 bestimmt

Personen unscharf abgebildet wurden (siehe Abb. 4.11). Eine der vorderen Personen des untersuchten Experiments ist zügig mit 1,4 m/s gegangen und wurde daher nicht erfasst.

PeTrack bietet die Möglichkeit, die inneren Strukturelemente zu ignorieren, und eine Person auch nur über die gesamte Kopfbedeckung bzw. über die äußere Kopfellipse (siehe Abschnitt 3.2.1) zu lokalisieren. Beim untersuchten Experiment hat dies zur Folge, dass keine falsch negative Detektion auftritt und damit die Falsch-Ermittlungs-Rate auf Null sinkt. Diese Einstellung hat aber zur Folge, dass immer dann, wenn eine Person temporär nicht über die Markerelemente gefunden wird, die gesamte Kopfbedeckung zur Detektion herangezogen und der Mittelpunkt der approximierenden Ellipse als Position genommen wird, was zu einer ungenaueren Lokalisierung führt, da eine temporäre reine Verfolgung eines Strukturelementes robust und genau ist (siehe Abschnitt 3.3).

Das Auslesen des Farbmarkerelementes (siehe Abschnitt 3.2.1) war bei detektierten Markerelementen robust. Lediglich bei der schwarzen Farbe als Größenindikator und daraus resultierender Nutzung des Umfangs des eingeschlossenen dunklen Bereichs, kam es manchmal zur Verwechslung des Farbmarkerelementes mit dem schwarzen Punkt als drittes Markerelement (siehe Abb. 3.32). Für die Ermittlung der Farbe zur Größenbestimmung führte dies zu keinem Fehler, da beide Elemente schwarz waren. Jedoch diente die Lage des Farbpunktes beim Kasernen- und Japan-Marker auch zur Bestimmung der Kopfausrichtung, so dass eine Fehlzuordnung eine angenommene Kopfdrehung von 180° zur Folge hätte.

Sollte nun die Kopfausrichtung untersucht werden, wurde daher eine Markererkennung in jedem Frame durchgeführt und nicht die zur Zeitersparnis implementierte Möglichkeit genutzt, die Erkennung nur gelegentlich durchzuführen und dazwischen die Personen lediglich zu verfolgen. Sind nun in jedem Frame die Personen detektiert worden, können die seltenen 180°-Drehungen deshalb ausgeschlossen werden, da diese nicht innerhalb von 0,04 s vollzogen werden können (die schnelle mögliche Änderung der Blickrichtung ist nicht identisch mit der Kopfausrichtung). In PeTrack wurde dieser Fehler dadurch eliminiert, dass die Kopfausrichtung mit benachbarten Frames verglichen wurde und bei einer 180°-Drehung die Markerelemente des aktuellen Frames getauscht wurden.

Beim Japan-Marker kam es zu keinem Fehler bei der Detektion und auch die Unterscheidungsmöglichkeit der Markerelemente (schwarzer und roter Punkt) zur Bestimmung der Kopfausrichtung war immer gegeben (siehe Abb. 3.8 3).

Die beiden in Tab. 3.4 untersuchten Korridor-Versuche wurden in einem geschlossenen System durch-

Dichte [1/m ²]	erkannte Personen	falsch positiv	falsch negativ
0,5	92	0% (0)	0% (0)
1,0	114	2,6% (3)	0% (0)
1,5	87	4,6% (4)	1,2% (1)
2,5	122	3,3% (4)	0% (0)
4,0	79	3,8% (3)	0% (0)

Tab. 3.5: Dichteabhängige Falsch-Ermittlungs- und Falsch-Negativ-Raten bei geschlossenen unidirektionalen Korridor-Versuchen der Versuchsreihe Nr. 6 gemäß Tab. 4.1 unter Nutzung des Hermes-Markers (Nr. 2); es wurden jeweils die ersten 1000 Frames ohne Berücksichtigung des Höhenfeldes oder einer Hintergrundsubtraktion untersucht

geführt, um die globale Personendichte exakt einstellen zu können. Sie unterscheiden sich dahingehend, dass bei den Kasernen-Versuchen (Nr. 3) nur eine eindimensionale Bewegung möglich war (siehe Abb. 4.2 a), hingegen bei den Hermes-Versuchen (Nr. 6) die Breite des Korridors 180 cm betrug und damit eine zweidimensionale Dynamik existierte (siehe Abb. 4.3 b). Wie bei der Einführung der Marker auf Seite 57 bereits erwähnt, ist der Marker 3 am fehleranfälligsten, was auch der Tab. 3.4 zu entnehmen ist.

Für den geschlossenen Korridor-Versuch der Versuchsreihe Nr. 6 ist in Tab. 3.5 aufgeführt, wie die Anzahl falsch positiver und negativer Detektionen mit der Dichte zusammenhängen. Die Bestimmung der Laufwege wurde für alle Versuche in den ersten 1000 Frames durchgeführt ohne das Höhenfeld (siehe Abschnitt 3.6.6) oder die Hintergrundsubtraktion (siehe Abschnitt 3.5.1) zur Verringerung des Fehlers heranzuziehen. Außer bei der geringsten Dichte von 0,5/m² bewegt sich der Fehler in ähnlichen Größenordnungen mit einer Falsch-Positiv-Rate von ca. 3,5% und einer nahezu verschwindenden Falsch-Ermittlungs-Rate. Fast alle falsch positiven Detektionen rühren von ähnlichen Strukturen wie dem Marker her, die am Hinterkopf der Personen entstehen, wenn die Schnüre der Bandana Kopfbedeckung mit dunklen Haaren der Versuchsteilnehmer auch einen weiß umrandeten schwarzen Fleck beschreiben (siehe Abb. 3.69). Einige dieser Fehler konnten darüber abgefangen werden, dass die fehlerhaften Punkte nahe des eigentlichen Markers liegen und bei einer aktivierten Verschmelzung von nahen Laufwegen mit der entsprechenden richtigen Trajektorie zusammengeführt werden. Dass bei der geringsten Dichte kein Fehler auftrat, kann damit erklärt werden, dass die Personen ungehindert gehen konnten und bei der freien Gehgeschwindigkeit den Blick nach vorne richten und damit die Fehler verursachenden Schnüre am Hinterkopf nicht sichtbar waren. Die falsch negative Detektion wurde durch eine sehr schief aufgesetzte Kopfbedeckung ähnlich der oberen Person in Abb. 3.22 hervorgerufen.

Bei der Bestimmung der Laufwege für das in Tab. 3.4 aufgeführte Tribünen-Experiment wurden die



Abb. 3.69: Ähnliche Struktur (rechts) wie der Hermes-Marker (links) am Hinterkopf einer Person aufgrund von Schnüren der Bandana Kopfbedeckung und dunklen Haaren des Versuchsteilnehmers, was zu Fehldetektionen führen kann

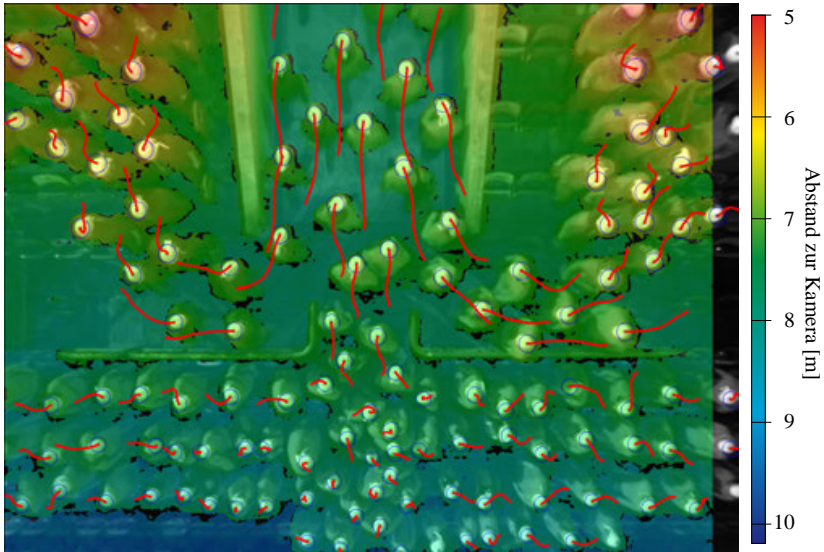


Abb. 3.70: Überkopfaufnahme einer Stereokamera aus Abb. 4.8 im Unterrang; die Einfärbung entspricht der Entfernung zur Kamera (rot = nah, blau = fern); die roten Pfade beschreiben den Laufweg jeder Person in der letzten Sekunde; siehe Video Nr. 44 für den gesamten Versuchsdurchlauf; ähnliche Bilder wurden verwendet in [506–508]

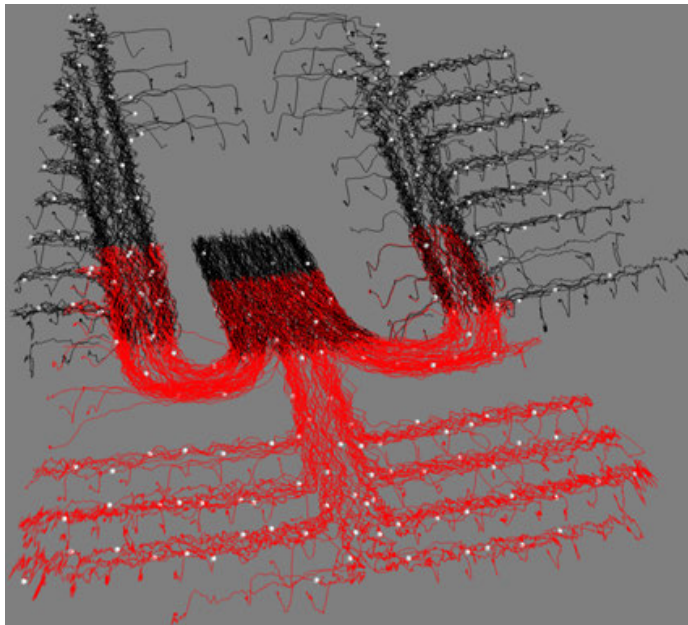


Abb. 3.71: Alle erfassten Laufwege des Versuchs aus Abb. 3.70, wobei die schwarzen Pfade von der oberen und die roten von der unteren Kamera erfasst wurden; die weißen Kugeln zeigen die Personenverteilung zum Zeitpunkt der Aufnahme in Abb. 3.70; siehe Video Nr. 45 für den gesamten Versuchsdurchlauf

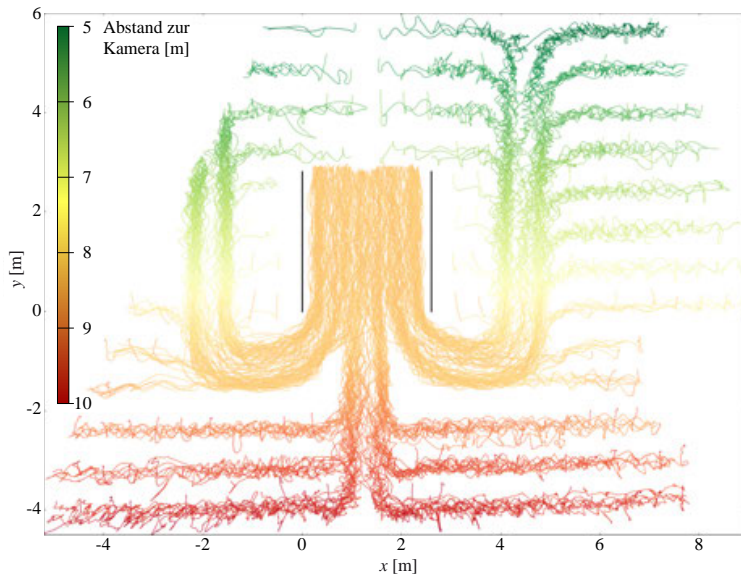


Abb. 3.72: Aufsicht auf alle ermittelten Laufwege eines Versuches im Tribünen Oberrang aus Versuchsreihe Nr. 6 (siehe auch Abb. 4.8); die Einfärbung entspricht der Entfernung zur Kamera (rot = fern, grün = nah)

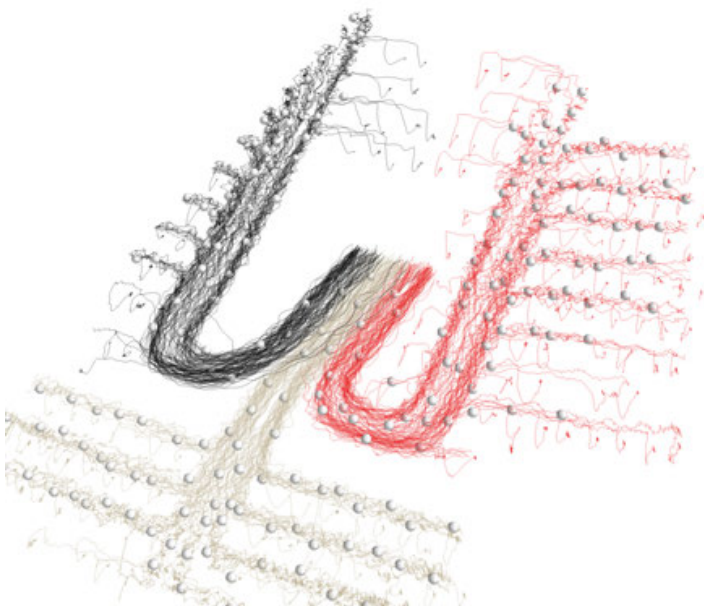


Abb. 3.73: Alle erfassten Laufwege des Versuchs aus Abb. 3.70 eingefärbt entsprechend der Shirtfarbe (siehe Abb. 4.8); die weißen Kugeln zeigen die Personenverteilung zum Zeitpunkt der Aufnahme in Abb. 3.70; siehe Video Nr. 48 für den gesamten Versuchsdurchlauf

Einstellungen wie im vorher betrachteten Versuch vorgenommen: ohne Nutzung des Höhenfeldes der Stereokamera, keine Hintergrundsubtraktion und Verschmelzung von nahen Laufwegen. Alle zu viel erkannten Personen rühren von Problemen beim Tragen der Kopfbedeckung her (siehe Abb. 3.69) und ereigneten sich im unteren Teil der Tribüne, in dem die Personen anfangs mit dem Hinterkopf zur Kamera saßen. Die Anzahl falsch positiver Detektionen wäre noch höher, wenn nahe Trajektorien nicht zusammengeführt worden wären, was aber zu kleinen Sprüngen im Laufweg führen kann. Die einfache Struktur des Markers führte zu Fehldetektionen, wobei diese selten in der komplexen Umgebung der Tribüne gefunden wurde.

Ist aufgrund einer geringen Pixelabdeckung des Kopfes und einer zu erwartenden hohen Dichte im Personenstrom wieder ein einfach strukturierter Marker nötig, so wäre in künftigen Versuchen die Hinzunahme eines weiteren Merkmals in Erwägung zu ziehen. Dieses Merkmal könnte ein markant gefärbter Punkt anstatt eines schwarzen Punktes sein.

Der Betrag der gemäß Gleichung 3.44 bestimmten mikroskopischen Beschleunigung in Tab. 3.4 zeigt, dass dieser durch die exakte Positionsbestimmung gering ist. Er ist zudem auch geringer als beim Experiment an der T-Kreuzung aus Abschnitt 3.6.6, da hier die Experimente mit maximaler Dichte und damit der geringsten Geschwindigkeit aufgeführt sind. Beim Experiment in der Tribüne mit der am Abstand größten Beschleunigung von $2,2\text{ m/s}^2 \pm 2,7\text{ m/s}^2$ wurde für einen passenden Vergleich auch nur die Bewegung in der Ebene herangezogen, so dass sich die vertikale Bewegung auf den Stufen nicht auf das Ergebnis auswirkt. Wohl aber die Art der Positionsbestimmung. Wurde für die ebenen Versuche eine durchschnittliche Personengröße je Person entlang ihres Laufweges ermittelt und diese zur Positionsbestimmung herangezogen, so floss die aufgrund der Disparität bestimmte Kameraentfernung direkt in die Gleichung 3.18 zur Ermittlung der Position ein. Hierdurch hat die vertikale Bewegung wie auch die Ungenauigkeit in der Bestimmung der Disparität direkten Einfluss auf die Position in der betrachteten horizontalen Bewegungsebene. Die Ungenauigkeit in der Entfernungsermittlung ist bei großem Abstand zur Kamera entsprechend Gleichung 3.25 bzw. 3.26 größer, so dass im unteren Tribünenbereich die Schwankung in den Trajektorien am stärksten ist.

Visualisierte Laufwege aus Tribünen-Versuchen sind Abb. 3.70, 3.71, 3.72 und 3.73 zu entnehmen. Dort ist auch der unruhigere Verlauf entfernter Trajektorien zu erkennen. Ein synchronisiertes Video der Ansichten ist Video Nr. 45 zu entnehmen. Einen weiteren Zusammenschnitt mit einer entsprechenden Simulation zeigt Video Nr. 43. Die Geschwindigkeit entlang der Laufwege wird in Video Nr. 47 dargestellt. Jede Visualisierung betrachtet einen anderen Schwerpunkt: Abb. 3.70 ist das Bild einer Perspektive der unteren Stereokamera mit überlagerten Daten der Extraktion, Abb. 3.71 veranschaulicht, aus welchen beiden Trajektoriensätzen das Gesamtergebn zusammengesetzt wurde (vgl. Abschnitt 3.3.5), Abb. 3.72 ist die orthogonale Projektion der Laufwege auf die horizontale Bewegungsebene, die bei der Berechnung der Beschleunigung verwandt wurde und Abb. 3.72 zeigt durch Einfärbung der Laufwege entsprechend der Shirtfarbe bzw. des Sitzplatzes, dass u. a. die Durchmischung der drei Ströme von den unterschiedlichen Tribünenbereichen innerhalb des Mundloches gering ist.

3.6.4 Trajektorien mit Farbmarkern markierter Personen

Mit Farbmarkern in Form einer farbigen Mütze konnte bei dem bisher einzigen Einsatz in Versuchsreihe Nr. 8 für alle Versuche eine fehlerfreie Erkennung durchgeführt werden. Beispiele dieser Versuche sind Abb. 3.9 4, 3.15, 3.76 und 4.3 e zu entnehmen.

3.6.4.1 Parameterwahl

Mit den in Abschnitt 3.2.2 vorgestellten Parametern wie

- der minimal und maximal zulässigen Anzahl an zusammenhängenden Pixeln des gesuchten Farbbereichs,
- dem maximalen Seitenverhältnis der approximierenden Ellipsen,
- dem dreidimensionalen Farbintervall im HSV-Farbraum und
- dem möglichen Öffnen und Schließen der Maske

konnte jederzeit eine Separation der Personen voneinander erreicht werden. Die Parameter mussten zwischen den Kameraeinstellungen geringfügig angepasst werden, je nachdem welche Optik und Brennweite verwandt wurde. Das Farbintervall musste insbesondere darauf abgestimmt werden, ob ein Versuch in der Sonne oder im Schatten durchgeführt wurde. Die zulässige Pixelanzahl wurde zwischen dem 0,5- und 1,2-fachen der maximal möglichen Anzahl eingestellt. Diese lag z. B. für Versuche entsprechend Abb. 3.76 bei

$$\pi(14\text{ cm} \frac{w/2}{z_0 \tan \alpha})(21\text{ cm} \frac{w/2}{z_0 \tan \alpha}) \approx 165\text{ px}$$

bei Annahme einer elliptischen Abbildung des Kopfes mit einer durchschnittlichen Ausdehnung von $14\text{ cm} \times 21\text{ cm}$, einem horizontalen Öffnungswinkel der Kamera von $2\alpha = 69^\circ$, einer horizontalen Bildauflösung von $w = 1920\text{ px}$, einer Aufhängehöhe von 2035 cm und einer Durchschnittspersonengröße von 173 cm ($z_0 = 2035\text{ cm} - 173\text{ cm} = 1862\text{ cm}$). Die große Spanne ist für die kleine Intervallgrenze insbesondere der engen Eingrenzung des Farbintervalls zur Vermeidung von Fehldetektionen und für die große Intervallgrenze der unterschiedlichen Kopfgrößen, der idealisierten Annahme der elliptischen Abbildung und der möglichen Ausweitung durch Haare und Gesicht geschuldet. Eine diffuse Beleuchtung erlaubt hier wie bei allen Markern eine robustere Detektion. Das maximale Seitenverhältnis der approximierenden Ellipsen wurde auf 2 gesetzt und liegt damit zur robusteren Detektion über dem idealen Seitenverhältnis von $21/14 = 1,5$. Das Farbintervall entsprach dem Intervall für Abb. 3.15, das ausführlich in Abschnitt 3.2.2 besprochen wird. Für die Größe des strukturierenden Elementes zum Schließen und Öffnen wurde $1/4$ der Kopflänge gewählt:

$$\frac{21}{4}\text{ cm} \frac{w/2}{z_0 \tan \alpha} \approx 4\text{ px}.$$

Zu einer Verschmelzung der Marker kam es nicht, da bei Versuchen, die auch am Rand der Kamera durchgeführt wurden, die Dichte für eine gegenseitige Verdeckung zu gering war, und bei hohen Dichten die Experimente unter einem geringen Blickwinkel stattfanden (siehe Abb. 3.9 4 und 3.15). Der geringe Blickwinkel auf die kompletten Versuchsbereiche konnte dadurch erzielt werden, dass die Kameraaufhängung in 2035 cm über dem Boden erfolgte.

Abb. 3.74 zeigt, dass trotz kritischer Farben im Bild die Parameter des Farbmarkers ausreichen, um genau nur die farbigen Mützen nicht zu maskieren. Weißtöne im Bild sind immer ungünstig, da sie im HSV-Farbraum bei Verringerung der Sättigung der Farbe z. B. hervorgerufen durch Überbelichtung von Teilen der Aufnahme und leichter Verschiebung der Farbtemperatur nah an jedem Farbton im

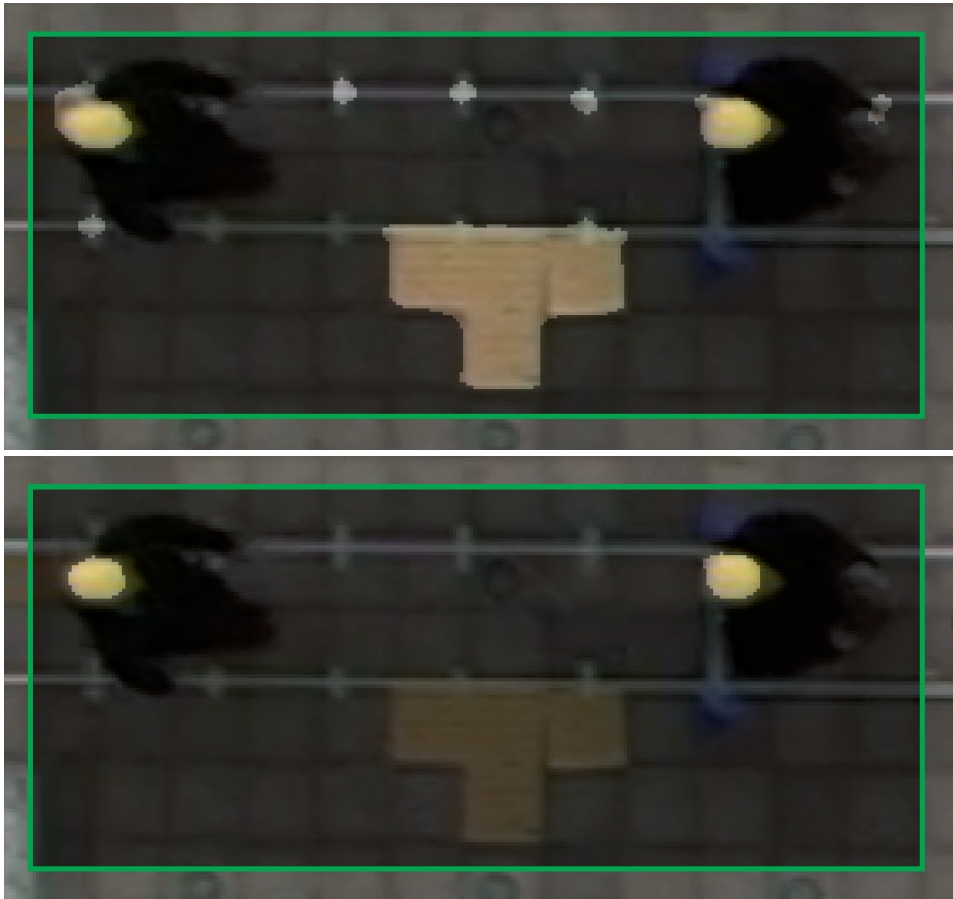


Abb. 3.74: Durch Anpassung der Parameter des Farbmarkers können auch kritische Bereiche meist zuverlässig ausgeblendet werden; die untere Maske entsteht aus der oberen durch leichte Änderung des Farbbereichs und einen größeren Durchmesser des strukturierenden Elementes beim Öffnen und Schließen

ungesättigten Bereich liegen. Die vier gelblichen Bodenplatten liegen farblich nah an der Mützenfarbe. Die Anpassung des Farbbereichs und der Größe des strukturierenden Elementes zum Schließen und Öffnen erlaubt es, die falsche Maske des oberen Bereiches der Abb. 3.74 zu verbessern, so dass die untere Maske erzielt wird. Aber auch durch die oben aufgeführte Wahl der minimalen und maximalen Pixelanzahl sowie des Seitenverhältnisses würden nur die Personen detektiert werden. Die Bodenfliesen könnten auch dadurch aus der Auswahl entfernt werden, indem der interessierende grün eingegrenzte Bereich (ROI) angepasst würde.

3.6.4.2 Fehler durch perspektivische Sicht

Durch die perspektivische Sicht auf die Köpfe der Personen ist von den als farbige Mützen getragenen Farbmarkern nur ein Teil zu sehen. Dieser Teil ist abhängig von dem Betrachtungswinkel α und der vertikalen Kopffentfernung z_0 zum optischen Zentrum.

Bei der Form des Kopfes gibt es viele, zum Teil auch regionaltypische Unterschiede [509]. Von oben und im oberen Bereich bis zu den Schläfen kann die Form eines Erwachsenen Kopfes durch einen Ellipsoid der Form

$$\frac{x^2}{r_x^2} + \frac{y^2}{r_y^2} + \frac{z^2}{r_z^2} = 1$$

mit den Längen der Halbachsen von $r_x = r_z = 7$ cm und $r_y = 10,5$ cm dargestellt werden (Anleihen aus [510, 511]).

Zur Vereinfachung betrachten wir die Wirkung der perspektivischen Projektion entlang der Frontalan-sicht und nehmen eine Kopfbedeckung an, die bis zur Schläfe eng und rundherum gleichmäßig anliegt (siehe gelbe Mütze in Abb. 3.75). Für andere Ansichten und Kopfbedeckungen ergeben sich vergleichbare Berechnungen. Für diese Ansicht vereinfacht sich die Form des Kopfes zu einem Halbkreis

$$y = \sqrt{r^2 - x^2}, \quad x \in [-r, r]$$

mit dem Radius $r = 7$ cm und der Steigung

$$y' = -\frac{x}{\sqrt{r^2 - x^2}}, \quad x \in]-r, r[.$$

Für die linke Kopfhälfte ($x \in [-r, 0]$) bzw. bei einer Person links vom Kameralot (rechte Kopfhälfte entsprechend mit anderem Vorzeichen) hängt der Blickwinkel $\beta \in [0, \pi/2]$ zum äußersten sichtbaren Punkt mit diesem wie folgt zusammen:

$$\begin{aligned} \cot \beta &= -\frac{x}{\sqrt{r^2 - x^2}} \Leftrightarrow x = -\frac{r}{\sqrt{1 + (\tan \beta)^2}} \\ &\Rightarrow x = -r \cos \beta. \end{aligned}$$

Der äußerste noch sichtbare Punkt in der x - z -Ebene ist somit $(-r \cos \beta \ 0 \ r \sin \beta)^T$.

Der Mittelpunkt der approximierenden Ellipse in der Projektionsebene liegt für Personen links vom Kameralot in Kopfhöhe für $x \geq r$ in der Mitte der gestrichelten gelben Linie in Abb. 3.75, deren Enden die Projektionen der Punkte $(-r \cos \beta \ 0 \ r \sin \beta)^T$ und $(r \ 0 \ 0)^T$ sind, die die äußersten Punkte der gelben Kopfbedeckung darstellen. $0 \leq x < r$ wird nicht weiter betrachtet, da für die nahezu senkrechte Betrachtung der Fehler e_c gering ist.

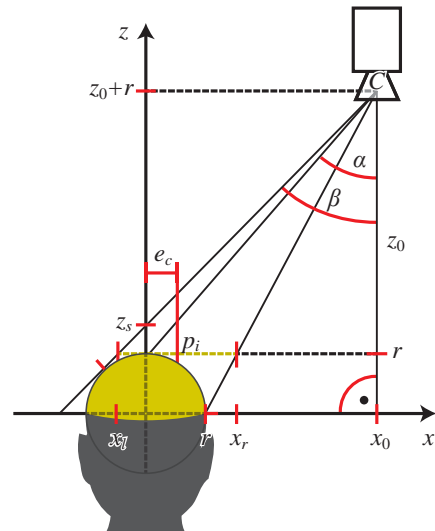


Abb. 3.75: Angenommene Lage des durch einen Halbkreis vom Radius r approximierten Farbmarkers in der x - z -Ebene bei einem Blickwinkel von α , der Lage des optischen Zentrums bei $C = (x_0 \ 0 \ z_0 + r)^T$ und der realen Position einer Person bei $(0 \ 0 \ z_0)^T$; die gestrichelte gelbe Linie im Intervall von $[x_l, x_r]$ gibt in Kopfhöhe mit dem Mittelpunkt den Fehler e_c an, der durch die perspektivische Sicht hervorgerufen wird

Das optische Zentrum liegt in der x - z -Ebene bei $(z_0 \tan \alpha \ 0 \ z_0 + r)^T$. Mit $x_0 = z_0 \tan \alpha$ liegt die x -Koordinate des rechten Randes der Abbildung des Farbmarkers x_r gemäß Strahlensatz bei

$$\frac{x_0 - r}{x_0 - x_r} = \frac{z_0 + r}{z_0} \Leftrightarrow x_r = r \frac{z_0 + x_0}{z_0 + r}$$

und damit der projizierte Punkt bei $\left(r \frac{z_0 + x_0}{z_0 + r} \ 0 \ r\right)^T$.

Mit $\sin \beta = r/z_s$ und $\tan \beta = -x_l/(z_s - r)$ ergibt sich

$$x_l = \left(r - \frac{r}{\sin \beta}\right) \tan \beta = r \frac{\sin \beta - 1}{\cos \beta}. \quad (3.30)$$

β lässt sich durch x_0 und z_0 ausdrücken, so dass am Ende die Grenzen nur von zwei Parametern abhängen. Sei α' der Winkel zwischen der optischen Achse und der Geraden vom optischen Zentrum zum Ursprung, so gelten folgende Gleichungen:

$$\tan \alpha' = \frac{x_0}{z_0 + r}, \quad \sin(\beta - \alpha') = \frac{r}{\sqrt{x_0^2 + (z_0 + r)^2}}.$$

Daraus ergibt sich:

$$\beta = \arcsin \frac{r}{\sqrt{x_0^2 + (z_0 + r)^2}} + \arctan \frac{x_0}{z_0 + r}.$$

Eingesetzt in Gleichung 3.30 lässt sich x_l in Abhängigkeit von x_0 und z_0 schreiben:

$$x_l = r \frac{\sin \left(\arcsin \frac{r}{\sqrt{x_0^2 + (z_0 + r)^2}} + \arctan \frac{x_0}{z_0 + r} \right) - 1}{\cos \left(\arcsin \frac{r}{\sqrt{x_0^2 + (z_0 + r)^2}} + \arctan \frac{x_0}{z_0 + r} \right)}.$$

Der Mittelpunkt der detektierten Farbfläche liegt bei $((x_r + x_l)/2 \ 0 \ r)^T$ und damit der beim Farbmarker als Position einer Person i festgestellte Aufenthaltsort im Koordinatensystem von Abb. 3.75 bei

$$p_i = \left(\frac{r}{2} \left(\frac{z_0 + x_0}{z_0 + r} + \frac{\sin \left(\arcsin \frac{r}{\sqrt{x_0^2 + (z_0 + r)^2}} + \arctan \frac{x_0}{z_0 + r} \right) - 1}{\cos \left(\arcsin \frac{r}{\sqrt{x_0^2 + (z_0 + r)^2}} + \arctan \frac{x_0}{z_0 + r} \right)} \right) \ 0 \ r \right)^T. \quad (3.31)$$

Der Fehler e_c aufgrund der perspektivischen Sicht kann direkt der x -Koordinate von p_i entnommen werden und lautet in der anfänglich beschriebenen Abhängigkeit von $\alpha = \arctan(x_0/z_0)$ und z_0

$$e_c = \frac{r}{2} \left(\frac{z_0 + z_0 \tan \alpha}{z_0 + r} + \frac{\sin \left(\arcsin \frac{r}{\sqrt{(z_0 \tan \alpha)^2 + (z_0 + r)^2}} + \arctan \frac{z_0 \tan \alpha}{z_0 + r} \right) - 1}{\cos \left(\arcsin \frac{r}{\sqrt{(z_0 \tan \alpha)^2 + (z_0 + r)^2}} + \arctan \frac{z_0 \tan \alpha}{z_0 + r} \right)} \right). \quad (3.32)$$

Für Versuche aus Versuchsreihe Nr. 8, bei denen der Farbmarker genutzt wurde und die Probanden durch das ganze Bild gelaufen sind ergibt sich bei einer Aufhängehöhe von 2035 cm, einer Durchschnittspersonengröße von 173 cm und einem Weitwinkelkonverter (siehe Abb. 4.14 e) mit einem maximalen Öffnungswinkel von 69° ($\alpha = 34,5^\circ$) ein Fehler von $e_c \approx 4$ cm.

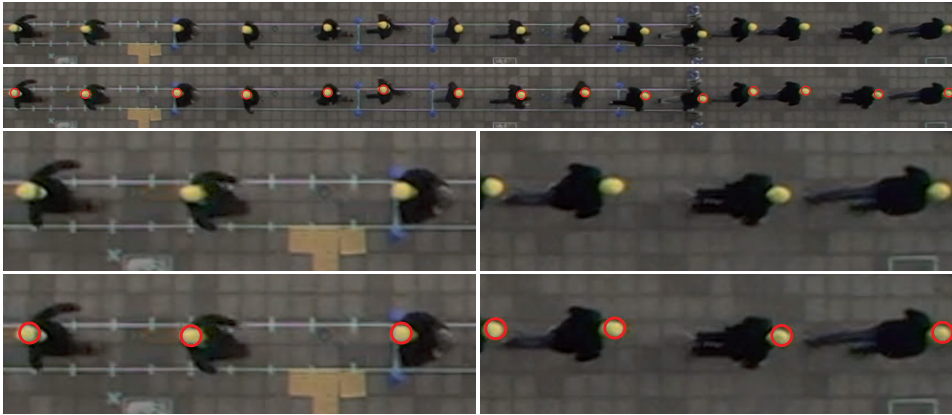


Abb. 3.76: Einsatz des Farbmarkers bei einem den ganzen horizontalen Öffnungswinkel ausnutzenden Versuch (siehe Video Nr. 27); der unterschiedliche Bereich der perspektivisch abgebildeten Mützen ist insbesondere in der unteren Vergrößerung zu erkennen; der Mittelpunkt der roten Kreise zeigt die ermittelte Position an

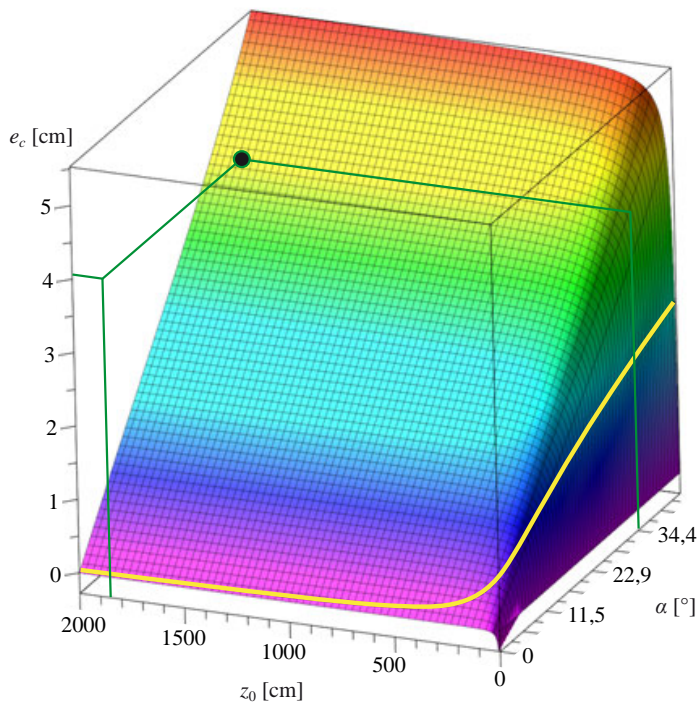


Abb. 3.77: Funktion des Fehlers e_c in Abhängigkeit vom Betrachtungswinkel α und der Entfernung zwischen Kamera und Kopf von z_0 , der von der perspektivischen Sicht auf den Farbmarker herrührt; unterhalb der gelben Kurve ist die Bedingung $a \geq \arctan(r/z_0)$ nicht erfüllt und damit Gleichung 3.32 nicht anwendbar; der abgebildete Bereich $z_0 \in [0\text{cm}, 2000\text{cm}]$ und $\alpha \in [0^\circ, 45^\circ]$ zeigt den typischen Wertebereich dieser Parameter; der Einfluss der Kameraentfernung auf den Fehler ist gering und der des Winkels annähernd linear; der schwarze Punkt markiert den maximalen Fehler am Bildrand für den Versuch aus Abb. 3.76

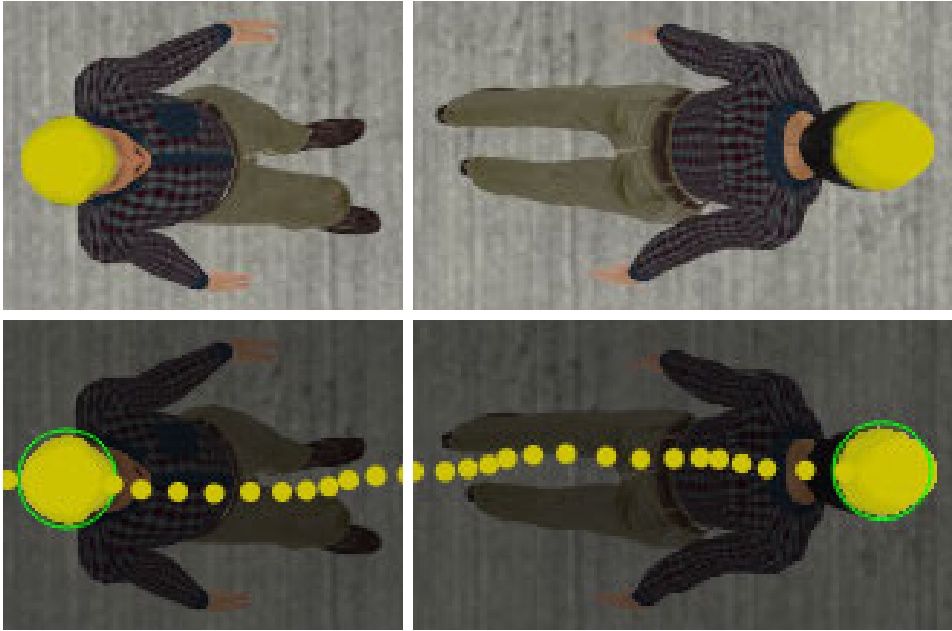


Abb. 3.78: Virtuelle Person mit eng anliegender farbiger Mütze als Farbmarker, abgebildet unter großen Blickwinkeln; (links) Blickwinkel entspricht Abb. 3.90 und Abb. 3.81; (unten) farbige unpassende Pixel sind maskiert und der Laufweg eingezeichnet; der Mittelpunkt des grünen Kreises ist die aktuell ermittelte Position

Abb. 3.76 zeigt einen Versuch dieser Versuchsreihe, bei dem im linken Bildbereich die farbige Mütze von vorne und im rechten Bildbereich die Mütze von hinten detektiert wird. Der Mittelpunkt der roten Kreise ist jeweils die angenommene Position als Mittelpunkt der approximierenden Ellipse.

Der bei diesem Versuch gemäß Gleichung 3.32 gemachte Fehler ist in Abb. 3.77 als schwarzer Punkt markiert. Diese Abbildung zeigt den Verlauf des Fehlers in Abhängigkeit vom Betrachtungswinkel α und der Entfernung zwischen Kamera und Kopf von z_0 . Der Bereich unterhalb der gelben Kurve genügt nicht der Bedingung $a \geq \arctan(r/z_0)$, die bei der Herleitung von Gleichung 3.32 gefordert wurde. Unterhalb der Kurve würde der Fehler noch schneller als in Abb. 3.77 abfallen und ist damit für Personen nahe der optischen Achse vernachlässigbar. Der Funktionsverlauf zeigt gut, dass die Kameraentfernung für Abstände größer $z_0 = 200\text{cm}$ eine geringe Auswirkung auf den Fehler hat. Der Fehler e_c steigt hingegen mit dem Betrachtungswinkel α annähernd linear an:

$$e_c \approx 0,12\text{cm}/^\circ \alpha. \quad (3.33)$$

3.6.4.3 Fehlerbetrachtung anhand virtueller Person

Da von realen Versuchen keine Grundwahrheit der Position exakt ermittelt werden kann, soll der Fehler durch die perspektivische Sicht an einer virtuellen Person bestimmt werden. Hierzu geht wie in Abschnitt 3.6.5 und 3.6.6 eine Person horizontal mittig unter einer Kamera hindurch. Dabei sind alle Ein-

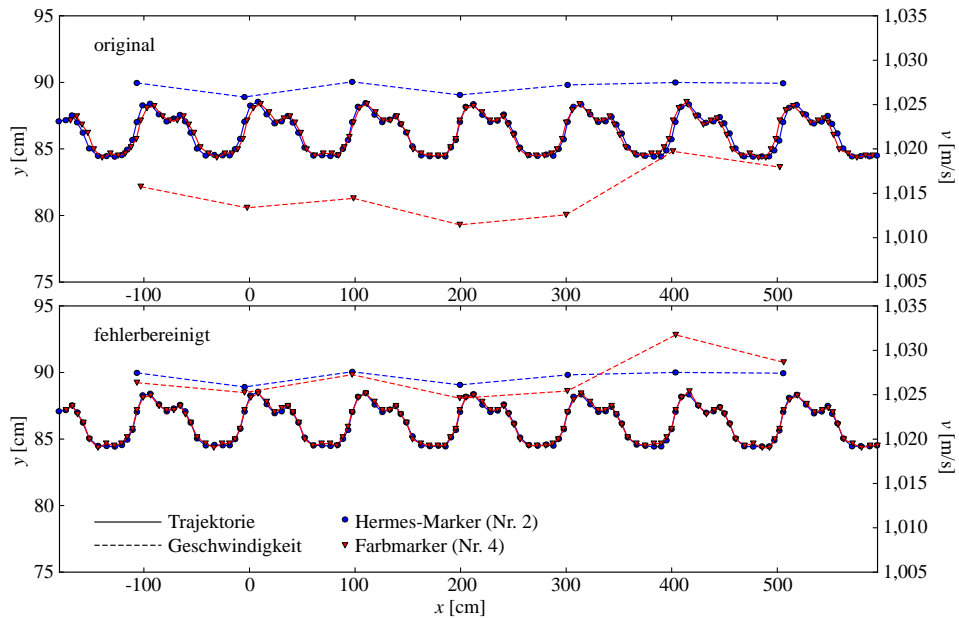


Abb. 3.79: Trajektorie und über eine Doppelschrittlänge gemittelte Geschwindigkeit derselben virtuellen Person, die über den Hermes- (Nr. 2) und Farbmarker (Nr. 4) erkannt wurde; (oben) sind die bis auf eine Gesamtverschiebung von $(3,6\text{cm } 0,5\text{cm})^T$ erhaltenen Originalpositionen und (unten) der für den Farbmarker gemäß Gleichung 3.34 angepasste Laufweg und die resultierende Geschwindigkeit entsprechend Gleichung 3.35 abgebildet

stellungen (u. a. Aufhängehöhe, Brennweite, Bildauflösung, Bildwiederholrate, Durchschnittspersonengröße) wie bei den ebenen Hermes-Versuchen gewählt worden (genauerer siehe Abschnitt 3.6.5).

Abb. 3.78 zeigt die eine gelbe Mütze tragende virtuelle Person zu zwei Zeitpunkten, in denen sie unter einem großen Blickwinkel gesehen wird. Die linke Seite entspricht dem Zeitschritt in Abb. 3.90 und Abb. 3.81. Im unteren Teil der Abbildung ist die nicht zum Marker passende Farbe maskiert und die erkannten Positionen während des Hindurchlaufens eingezeichnet. Der Mittelpunkt des grünen Kreises ist der Mittelpunkt der approximierenden Ellipse und zeigt damit die aktuell ermittelte Position an.

Der resultierende Laufweg ist in Abb. 3.79 oben im Vergleich zum Laufweg, der über den Hermes-Marker bestimmt wurde, abgebildet. Der Hermes-Marker ermöglicht es insbesondere bei der virtuellen Person, die exakte Position des Kopfes während des gesamten Laufweges zu ermitteln, da der Marker immer ganz zu sehen ist, parallel zur Bildebene liegt und der Kopf keine andere als die Laufbewegung vollzieht (siehe Abb. 3.90). Eine Gesamtverschiebung um $(3,6\text{cm } 0,5\text{cm})^T$ des exakten Laufweges war für den Vergleich nötig, da der Hermes-Marker nicht ganz kopfmittig aufgebracht war.

Der theoretisch bestimmte Fehler ist qualitativ zu beobachten, da die rote Trajektorie zur Mitte hin gestaucht erscheint. Dass auch für diesen Mützentyp der in Gleichung 3.33 angegebene quantitative Wert des approximierten Fehlers den realen Fehler gut abbildet, kann dem fehlerbereinigten Laufweg im unteren Bereich der Abb. 3.79 entnommen werden. Die erkannte Position liegt nach einer

Verschiebung gemäß Gleichung 3.33 bei

$$p_i^* = p_i + e_c = p_i + 0,12 \text{ cm} / ^\circ \begin{pmatrix} \alpha_x \\ \alpha_y \\ 1 \end{pmatrix} \quad (3.34)$$

und damit entlang des gesamten Laufweges sehr nah an der optimalen Position. Dabei sind α_x und α_y die vorzeichenbehafteten Blickwinkel zur optischen Achse entlang der jeweiligen Koordinatenachse.

Die Auswirkung des Fehlers und dessen Korrektur auf eine für die Fußgängerdynamik wichtige Kenngröße ist anhand der Geschwindigkeiten im Diagramm zu sehen. Durch die Positionsverschiebung wurde der Fehler in der Geschwindigkeit von bis zu 0,015 m/s (1,5%) nahezu eliminiert. Die Geschwindigkeit wurde unter Anwendung von Gleichung 4.10 mit einem Zeitschritt von $\Delta t = 1/16 \text{ s}$ gemäß

$$v_i(t) = \frac{1}{16} \sum_{j=0}^{15} \frac{\|p_i(t + (j+1)/16) - p_i(t + j/16)\|}{1/16 \text{ s}} = \frac{1}{s} \sum_{j=0}^{15} \left\| p_i \left(t + \frac{j+1}{16} \right) - p_i \left(t + \frac{j}{16} \right) \right\| \quad (3.35)$$

berechnet, um die mikroskopischen Geschwindigkeitsänderungen mit zu erfassen. Die Mittelung über 1 s bzw. 16 Frames und damit etwa über einen Doppelschritt war nötig, da der modellierte Bewegungsablauf der virtuellen Person nicht gleichmäßig ist, weshalb auch die Bestimmung der Glattheit der bestimmten Trajektorien der unterschiedlichen Verfahren an realen Daten erfolgen muss. Eine Glättung des Laufweges oder die alleinige Betrachtung der Hauptbewegungsrichtung wurde bei der Bestimmung der Geschwindigkeit, wie in Abschnitt 4.3 angegeben, nicht durchgeführt. Mit einer Einschränkung auf die Hauptbewegungsrichtung bei der Bestimmung der Geschwindigkeit liegt diese um ca. 0,005 m/s unterhalb der abgebildeten Werte.

Die Ähnlichkeit des Fehlers zu dem der markerlosen Detektion kann in Abschnitt 3.6.6 der Abb. 3.91 entnommen werden. Auch dort steigt der Fehler in der Trajektorie bei zunehmendem Blickwinkel.

3.6.4.4 Glattheit der Trajektorien

Der gemäß Gleichung 3.44 ermittelte Betrag der mikroskopischen Beschleunigung als Maß für die Glattheit (siehe Abschnitt 3.6.6) liegt für die mittels des Farbmarkers gewonnenen Trajektorien höher als bei Trajektorien, die mit Markern mit Strukturelementen erzielt wurden, da bei diesen immer genau das gleiche komplett im Bild sichtbare Element gesucht und als Position ausgewählt wird. Beim Farbmarker kommt es durch das Material der Mütze und einer durch den unterschiedlichen Lichteinfall variierenden Farbe zu weniger stabilen Detektionsorten. Diese variieren leicht auf dem Kopf, je nachdem, welche Pixel im Randbereich aufgrund Ihrer Farbe als noch gültig angesehen werden. Die Beschleunigung lag bei den bisher mit dem Farbmarker durchgeführten Versuchen in Versuchsreihe Nr. 8 zwischen $2,2 \text{ m/s}^2 \pm 1,7 \text{ m/s}^2$ und $2,9 \text{ m/s}^2 \pm 1,7 \text{ m/s}^2$ (vgl. Abschnitt 3.6.10).

Um den Fehler aufgrund der immer nur teilweise einsehbaren Mütze und der ggf. wegen uneinheitlicher Farbwiedergabe nicht komplett erfassten Mütze zu verringern, könnte der Farbmarker verkleinert werden. Dabei strebt der Farbmarker mit den Eigenschaften gegen die Marker mit strukturierenden Elementen und verliert die Einfachheit und die durch die Größe auch in großer Entfernung leichte Erkennbarkeit.

3.6.5 Disparität zur Personenerkennung

Die markerlose Personendetektion (siehe Abschnitt 3.5) geschieht über das aus der Disparitätsmatrix abgeleitete Höhenfeld. Daher ist die Qualität der Disparitätsmatrix von entscheidender Bedeutung für die Erkennung.

In diesem Abschnitt soll das beste der integrierten Verfahren (siehe Abschnitt 3.4.2, Seite 91) und die für dieses Verfahren entsprechenden Parameter bestimmt werden.

Für das Maß zur Bestimmung des besten Verfahrens sind z. B. denkbar:

1. durchschnittlicher Fehler der Disparität bzw. der entsprechenden Höhe pro Pixel für das gesamte Bild oder im Bereich des Vordergrundes,
2. durchschnittliche Abweichung von erzeugten Isolinien oder Ellipsen vom optimalen Verlauf,
3. durchschnittliche Abweichung der ermittelten Position von der über ein korrektes Höhenfeld bestimmten Kopfposition,
4. durchschnittliche Abweichung der ermittelten Position von der realen Position der Kopfmitte,
5. Anzahl an gefundenen, fälschlich gefundenen (falsch positiven) und nicht gefundenen (falsch negativen) Personen.

Ziel des Gesamtverfahrens ist die möglichst genaue Bestimmung von Laufwegen von Personen über ihre Kopfposition. Daher könnte ein Maß gemäß Punkt 1 oder 2 nur indirekt eine Aussage über die Qualität der erzeugten Trajektorien machen. Punkt 4 würde zwar den realen Fehler der Trajektorie wiedergeben, aber auch den Fehler beinhalten, der dem Ansatz über das perspektivische Höhenfeld und dem Bilden von Ellipsenpyramiden geschuldet ist und somit nicht dem Verfahren zur Disparitätsbestimmung anzulasten ist.

Aus diesem Grund soll im ersten Schritt der Fehler über die Abweichung gemäß Punkt 3 bestimmt werden. Da aber auch der reale Fehler interessant ist, wird dieser in Abschnitt 3.6.6 ebenfalls untersucht, um das Gesamtverfahren zu beurteilen. Punkt 5 ist dann von Bedeutung, wenn es Fehler dieser Art gibt.

Sei $p_i^*(t)$ die gemäß des Verfahrens aus Abschnitt 3.5 bestimmte optimale Position von Person i zum Zeitpunkt t unter Verwendung eines korrekten Höhenfeldes, so ergibt sich folgender Fehler als Maß für die Qualität des ermittelten Laufweges p_i von Person i über den Mittelwert der Abstände:

$$e_{\text{Verfahren}} = \frac{1}{F} \sum_{t=t_i^{\text{in}}}^{t_i^{\text{out}}} \|p_i(t) - p_i^*(t)\|. \quad (3.36)$$

Dabei ist $F = (t_i^{\text{out}} - t_i^{\text{in}})/\Delta t$ die Anzahl an Frames mit einem Einzelbildabstand von Δt , in denen Person i mit dem ganzen Kopf in beiden Perspektiven des Stereobildpaares zu sehen ist. Zudem bleibt der Bereich einer Bildfolge unbewertet, für den bei Verfahren der Klasse I aufgrund der Vergleichsfenstergröße $2\omega + 1$ (siehe Gleichung 3.21) keine Disparität bestimmt werden kann.

Da von einer realen Stereoaufnahme keine Grundwahrheit (*ground truth*) eines perspektivischen Höhenfeldes vorliegt, wurde eine virtuelle Szene mit der Software Panda3D [512] erstellt. Für einen besseren Vergleich entspricht die virtuelle Kamera den Randbedingungen der Stereokamera Nr. 4 (horizontaler Öffnungswinkel 66° (wg. Brennweite von 983 px bzw. 3,8 mm und Sensorgröße von $1/3''$),

Basislänge 24 cm, Bildwiederholrate 16 f/s, Auflösung $1280 \text{ px} \times 960 \text{ px}$) und die virtuelle Szene den ebenen Versuchen von Versuchsreihe Nr. 6 (Aufhängenhöhe 784 cm, Personengrößen $176 \text{ cm} \pm 9 \text{ cm}$).

Die virtuelle Szene hat zudem den großen Vorteil, das Höhenfeld unabhängig von möglichen Entzerrungsfehlern bestimmen zu können (siehe Abschnitt 3.6.2). Die Bestimmung geschah über pixelweises Raycasting vom optischen Zentrum der rechten Perspektive der virtuellen Stereokamera.

Zur Untersuchung der ermittelten Personenbewegung auf Basis des stark winkelabhängigen perspektivischen Höhenfeldes wurde eine über den gesamten Öffnungswinkel hinweg laufende Person in der Szene positioniert. Abb. 3.80 zeigt in den oberen beiden Reihen die mittig unter der virtuellen Stereokamera hindurch laufende Person. Dabei wurden für alle abgebildeten Schritte der relevante Bildausschnitt montiert, so dass die Person jede Sekunde einmal zu sehen ist. In der dritten (halbtransparente Überlagerung) und vierten Reihe ist das optimale Höhenfeld abgebildet. Die folgenden Reihen zeigen die Schritte des in Abschnitt 3.5 vorgestellten Verfahrens zur markerlosen Personenverfolgung: Isolinien äquidistanter Höhenschritte, approximierende Ellipsen und die Personenform beschreibende Ellipsenpyramide. Die letzte Reihe zeigt den kompletten über das optimale Höhenfeld ermittelten Laufweg. Hierbei ist zu erkennen, dass der Punkt, der die Position der Person angibt, auf dem Kopf wandert, und zwar in Richtung der Bildmitte (genauer siehe Abschnitt 3.6.6). Bei der am weitesten rechts stehenden nur halb abgebildeten Person verlaufen die Isolinien am gesamten Bildrand entlang und umschließen somit tieferliegende Bereiche, so dass keine gültigen Isolinien für die Personenerkennung erzeugt werden. Abb. 3.81 entspricht der zweiten Spalte aus Abb. 3.80. Die vollständige Sequenz zeigt Video Nr. 15.

Die gemäß vorgestelltem Verfahren optimale Position $p_i^*(t)$ konnte somit in jedem Frame bestimmt werden. Sie ist als Basis in Abb. 3.83 abgebildet. Mit den beiden in OpenCV implementierten Verfahren zur Bestimmung der Disparitätsmatrix wurde ebenfalls der Laufweg bestimmt. Dabei wurde der variable Parameter der Vergleichsfenstergröße $2\omega + 1$ über alle erlaubten Größen variiert. Der Fehler bzw. durchschnittliche Abstand $e_{\text{Verfahren}}$ gemäß Gleichung 3.36 und dessen Standardabweichung ist Tab. 3.6 für unterschiedliche Fenstergrößen zu entnehmen. Für das SGBM-Verfahren sind die erzeugten Disparitätsmatrizen mit überlagerten Höhenlinien im Abstand von 5 cm in Abb. 3.82 zu sehen und die Positionen in Abb. 3.83 eingetragen. Das ptGrey- und ELAS-Verfahren wurde nicht betrachtet, da, wie später zu sehen, diese Verfahren schlechtere Erkennungsraten liefern.

Der Tab. 3.6 ist zu entnehmen, dass die optimale Position am besten mit kleinen Fenstergrößen zu erzielen ist, wobei die Abweichung durchschnittlich 3 cm bei keiner Fenstergröße übersteigt. Die Standardabweichung des Fehlers wie auch die Beschleunigung korreliert kaum mit der Fenstergröße. Der Betrag der mikroskopischen Beschleunigung wurde gemäß Gleichung 3.44 zwischen jedem Einzelbild-Tripel bestimmt und sagt somit etwas über die Glattheit des ermittelten Laufweges aus. Dabei ist bereits die Beschleunigung des optimalen Laufweges mit $2,82 \text{ m/s}^2$ groß (vgl. Tab. 3.8), da die modellierte Gehbewegung der Person und damit das Schwingverhalten von der wechselseitigen Fußbelastung ggf. nicht optimal hinterlegt wurde.

Insbesondere der kleinen Vergleichsfenstergröße kommt bei der Bestimmung des Höhenfeldes und damit der Positionsbestimmung zu Gute, dass bei der Stereosequenz der virtuellen Szene

- kein Rest einer optischen Verzerrung vorliegt und somit gleiche 3D-Punkte auf jeden Fall genau auf der gleichen Pixelzeile liegen,
- Bilder optimal scharf abgebildet sind und
- die Textur auf dem Boden und am Oberkörper markant ist.

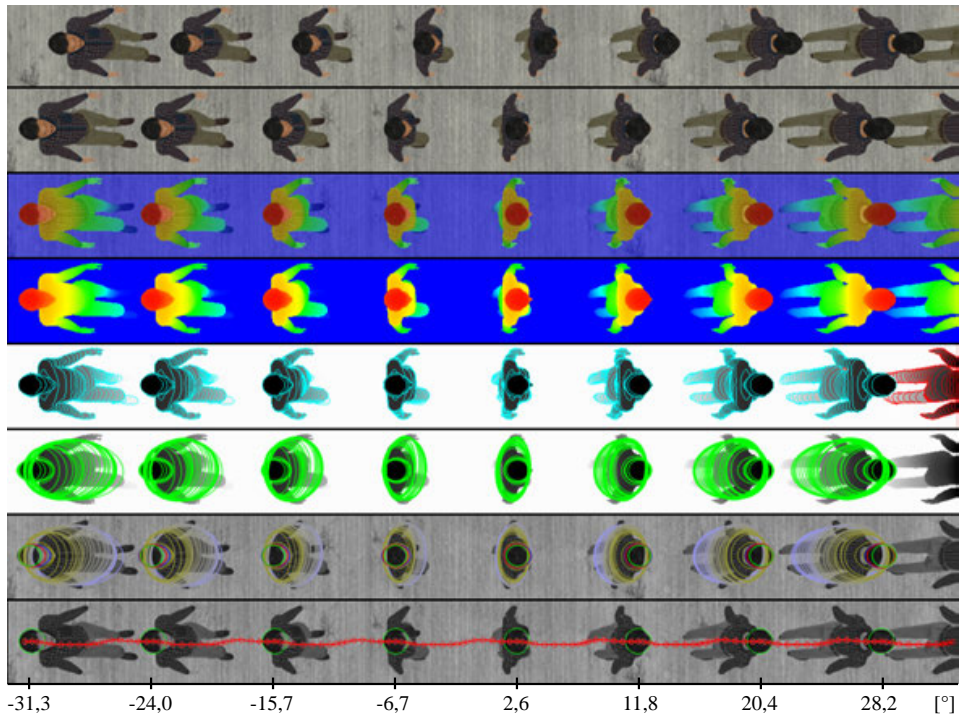


Abb. 3.80: Schritte der markerlosen Personenerkennung in einer virtuellen Szene mit exaktem Höhenfeld, bei der eine Person horizontal mittig unter einer Stereokamera hindurchläuft; von oben nach unten: linkes Stereobild, rechtes Stereobild, rechtes Stereobild überlagert vom exakten Höhenfeld, exaktes farbkodiertes Höhenfeld, Isolinien in 5-cm-Schritten, approximierende Ellipsen, Ellipsenpyramide, Laufweg; Bildmontage aus Bildsequenz im Sekunden-Abstand; vollständige Sequenz siehe Video Nr. 15; Ausschnitt der zweiten Spalte der Erkennungsschritte siehe Abb. 3.81; der Winkel ist der horizontale Blickwinkel zur optischen Achse

Für kleine Vergleichsfenstergrößen sind die optimalen Stereobilder wichtiger als für große Fenster, da sie empfindlicher auf Fehler unter anderem bei der Rektifizierung reagieren. Die häufigeren Fehlstellen in der Disparitätsmatrix bei kleiner Fenstergröße werden durch die lineare Interpolation ausgeglichen. Die Unterschiede der Disparitätsmatrizen in Abhängigkeit von der Fenstergröße ist gut anhand der Isolinien in Abb. 3.82 zu erkennen. Aus kleinen Fenstergrößen resultieren Disparitätsmatrizen, die unruhiger bzw. nicht so glatt verlaufen und mehr fehlende Disparitäten aufweisen, da die Korrespondenzsuche scheitert, aber das Höhenfeld exakter wiedergeben können. Dies ist u. a. an der größeren umrandeten obersten Höhenlinie zu erkennen, die damit am ehesten der optimalen Höhenlinie in Abb. 3.81 c entspricht. Die unruhigere Disparitätsmatrix führt zu entsprechend unruhigen Isolinien, die zudem häufiger kleine Ungenauigkeiten umschreiben, wie an der steigenden Zahl kleiner grüner Isolinien in Abb. 3.82 zu sehen ist. Den Unterschied von Disparitätsmatrizen, erzeugt mit kleiner und großer Fenstergröße, hat Ähnlichkeiten mit dem Unterschied zwischen einem Bild mit kleinen Pixelstörungen und einer Weichzeichnung dieses Bildes.

Nach den Fehlerwerten in Tab. 3.6 müsste entgegen der bisher in den Beispielen verwendeten großen Vergleichsfenster kleine genommen werden. Die Spalten zu den Falsch-Ermittlungs- und Falsch-Ne-

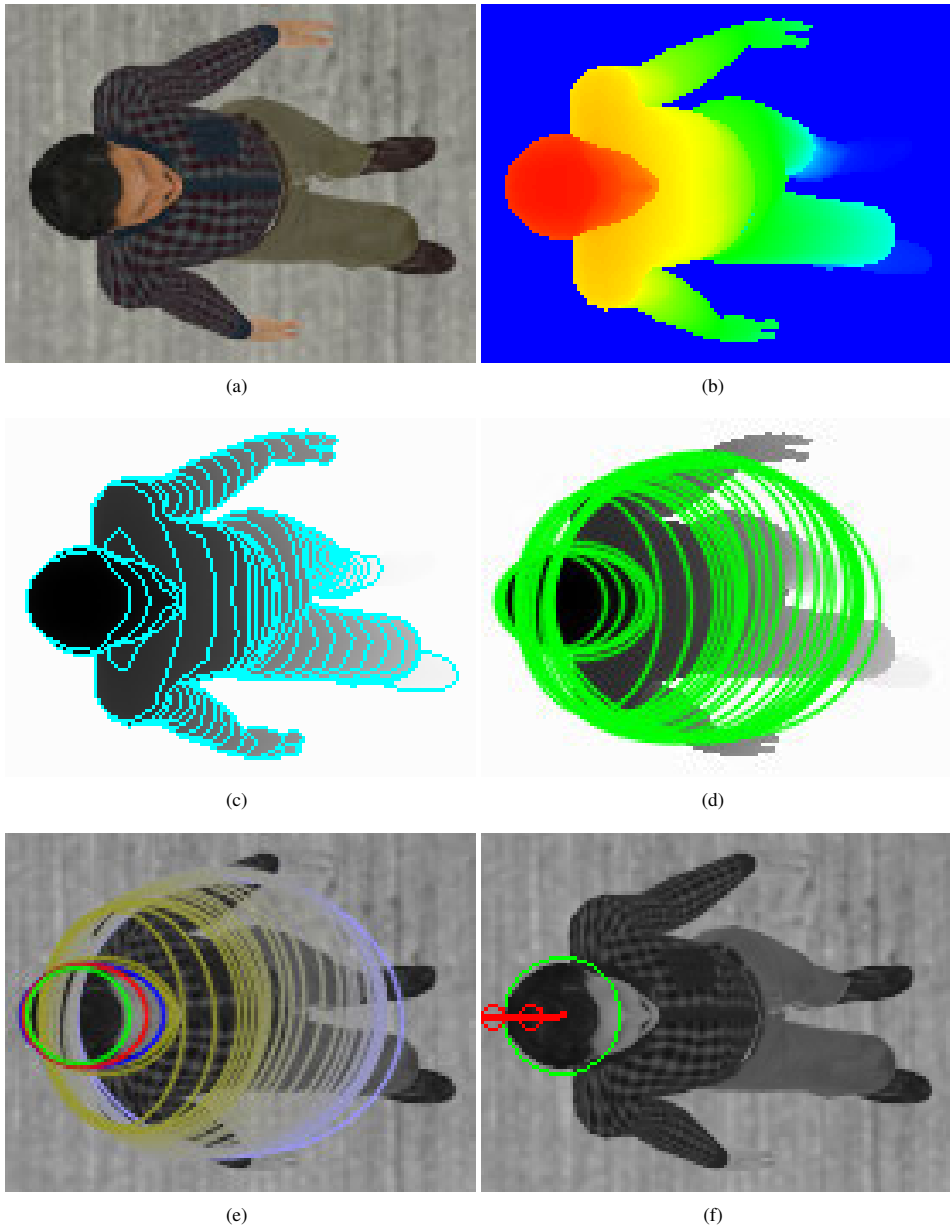


Abb. 3.81: Erkennungsschritte einer virtuellen Person auf Basis eines exakten Höhenfeldes: (a) Bild der rechten Perspektive der virtuellen Stereokamera; (b) exaktes farbkodiertes Höhenfeld; (c) Isolinien in 5-cm-Schritten; (d) die Isolinien approximierende Ellipsen; (e) Ellipsenpyramide; (f) ermittelte Kopfposition ist Mittelpunkt des grünen Kreises bzw. Ende des roten Laufweges; Ausschnitt der zweiten Spalte von Abb. 3.80

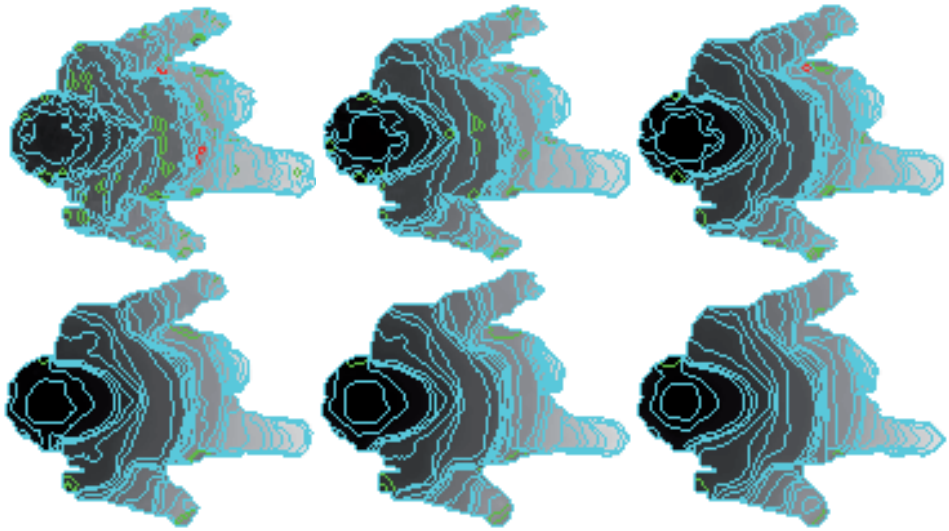


Abb. 3.82: Disparitätsmatrix mit überlagerten Höhenlinien im Abstand von 5 cm erzeugt unter Nutzung des SGBM-Verfahrens mit Vergleichsfenstergrößen von 1, 3, 5, 7, 9 und 11 px; Zeitpunkt und Ausschnitt entspricht der Aufnahme aus Abb. 3.81

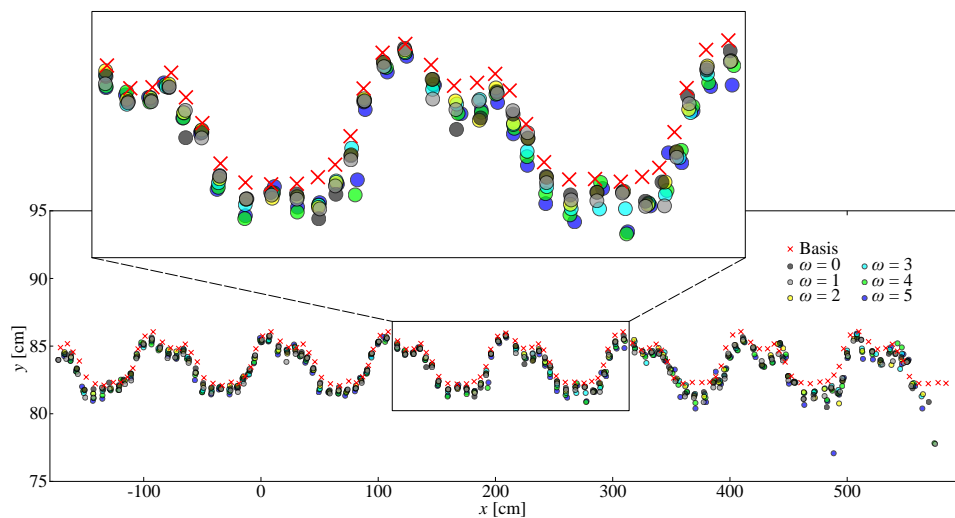


Abb. 3.83: Ermittelte Positionen durch markerlose Erkennung einer virtuellen Person mittels optimalem Höhenfeld als Basis und mittels Höhenfeld, das durch das SGBM-Verfahren mit unterschiedlichen Vergleichsfenstergrößen von $2\omega + 1$ erzeugt wurde

Verfahren	Fenstergröße [px]	Fehler $e_{\text{Verfahren}}$ [cm]	falsch positiv	falsch negativ	Beschleunigung [m/s ²]
Optimal		0	0%	0%	2,82 ± 1,51
BM	5	1,57 ± 1,19	0,9% (1)	0%	5,70 ± 4,39
	7	1,93 ± 1,94	0%	0%	6,85 ± 8,93
	9	1,92 ± 1,46	0%	0%	5,75 ± 4,86
	11	2,09 ± 1,80	0%	0%	6,30 ± 7,04
	13	2,09 ± 1,27	0%	0%	5,34 ± 3,88
	15	2,20 ± 1,34	0%	0%	5,33 ± 4,24
	17	2,33 ± 1,33	0%	0%	5,50 ± 4,29
	19	2,38 ± 1,22	0%	0%	5,10 ± 3,53
	21	2,33 ± 1,19	0%	0%	5,09 ± 3,05
SGBM	1	1,84 ± 1,63	0%	0%	4,48 ± 4,55
	3	1,68 ± 1,25	0%	0%	4,50 ± 3,80
	5	1,74 ± 0,98	0%	0%	3,89 ± 2,38
	7	2,11 ± 1,22	0%	0%	4,69 ± 3,59
	9	2,55 ± 1,63	0%	0%	5,16 ± 4,02
	11	2,85 ± 1,47	0%	0%	5,67 ± 3,89

Tab. 3.6: Für das BM- und SGBM-Verfahren sind für alle erlaubten Vergleichsfenstergrößen der Fehler bzw. durchschnittliche Abstand $e_{\text{Verfahren}}$ zur optimalen Position gemäß Gleichung 3.36, dessen Standardabweichung, die Falsch-Ermittlungs-Rate, die Falsch-Negativ-Rate und der Betrag der mikroskopischen Beschleunigung gemäß Gleichung 3.44 mit Standardabweichung angegeben; Fehldetektionen beziehen sich auf 117 untersuchte Frames mit einer Person

gativ-Raten sind in die Tabelle mit aufgenommen worden, um das Problem der kleinen Fenstergrößen aufzuzeigen. Bereits bei den optimalen Stereobildern der virtuellen Szene kommt es mit der kleinsten Fenstergröße zu einer falsch positiven Erkennung, also einer zu viel erkannten Person. Daher scheint es wichtig zu sein, auch das 5. oben aufgeführte Fehlermaß genauer zu untersuchen.

Um diesen Fehler quantifizieren zu können, wurde für alle integrierten Verfahren (gemäß Liste auf Seite 91) und deren freie Parameter (Vergleichsfenstergröße für die Korrespondenzsuche, Kantenmaskengröße beim Test auf Kantentreue für das ptGrey-Verfahren) getestet, wieviele der 1377 Personen in einem dichten Menschenstrom erkannt oder nicht erkannt (falsch negativ) und wieviele Personen zu viel (falsch positiv) erkannt wurden. Ein Hermes-Versuch (aus Versuchreihe Nr. 6) mit einer hohen Personendichte von ca. 1,7 Personen/m² und Wänden (ähnlich Abb. 4.18) wurde gewählt, um die kritischen hohen Dichten und die aufgrund des perspektivischen Höhenfeldes kritischen Wände (siehe Abschnitt 3.5.1) integriert zu haben.

In Tab. 3.7 sind die wesentlichen Ergebnisse zusammengefasst. Dabei bezieht sich die Fehlerrate auf den reinen Erkennungsschritt ohne anschließendes Tracking, so dass die Fehlerrate der erfassten Laufwege im Anschluss noch stark verringert werden kann (siehe Abschnitt 3.6.6). Die nicht erkannten Personen rühren meist von an der Wand laufenden Personen, die häufig auch mehrmals auf ihrem Weg nicht erkannt wurden. Die Falsch-Ermittlungs-Rate fällt bei steigender Vergleichsfenstergröße für alle Verfahren, lediglich beim ptGrey-Verfahren mit eingeschalteter Kantentreue und einer Maskengröße von 5 px liegt das Optimum bei der kleinsten Fenstergröße. Die Falsch-Negativ-Rate verhält

Verfahren	Fenster- größe [px]	Kanten- mg. [px]	falsch positiv	falsch negativ
ptGrey	5	-	32,7% (668)	1,1% (15)
		5	1,9% (26)	1,6% (23)
		11	9,6% (146)	1,9% (26)
	11	-	26,7% (501)	1,1% (15)
		5	9,0% (136)	3,4% (49)
		11	3,4% (49)	3,9% (56)
	21	-	15,7% (255)	8,5% (128)
		5	5,9% (87)	6,4% (95)
		11	2,5% (35)	7,9% (118)
BM	5		11,8% (184)	0,9% (13)
	13		5,2% (76)	1,2% (17)
	21		2,9% (41)	2,3% (32)
SGBM	1		2,3% (33)	1,6% (22)
	5		1,5% (21)	1,9% (26)
	11		0,7% (9)	2,9% (41)
ELAS			3,6% (51)	6,8% (101)

Tab. 3.7: Erkennungsfehler in Abhängigkeit von den integrierten Verfahren zur Disparitätsbestimmung und ihren Parametern; Spalten: Verfahrensnamen gemäß Liste auf Seite 91, Vergleichsfenstergröße für die Korrespondenzsuche, Kantenmaskengröße beim Test auf Kantentreue (nur für das ptGrey-Verfahren wählbar), Falsch-Ermittlungs- und Falsch-Negativ-Rate bei 1377 erkannten Personen in einem dichten (ca. 1,7 Personen/m² ähnlich Abb. 4.18) Personenstrom (reiner Erkennungsschritt ohne Tracking)

sich genau umgekehrt und steigt bei Erhöhung der Fenstergröße. Die beste Gesamtfehlerrate erzielt das SGBM-Verfahren mit einer Fenstergröße von 5 px, dicht gefolgt vom ptGrey-Verfahren mit Fenstergröße 5 px und Maskengröße 5 px. Beim ptGrey-Verfahren ist generell jedoch das resultierende Höhenfeld insbesondere zum Bildrand hin sehr störanfällig und erzielt außer mit der gerade aufgeführten Parameterkombination schlechte Ergebnisse. Auch das ELAS-Verfahren liefert schlechtere Erkennungsraten als die beiden Verfahren der OpenCV-Bibliothek.

Durch das anschließende Tracking ist die Falsch-Ermittlungs-Rate kritischer zu sehen, da derartige Fehlern, wenn sie im Vordergrund mehrmals an der gleichen Stelle auftreten (z. B. Schulter oder Rucksack) am Ende des Gesamtverfahrens Laufwege von Personen liefern, die gar nicht existieren. Hingegen kann die Falsch-Negativ-Rate sehr leicht durch das verbindende Tracking verringert werden, so dass ohne Vorgaben zur Erkennungsrate einer Person entlang der Trajektorie eine Person nur selten während ihres Weges unter der Kamera detektiert werden muss (im hier untersuchten Experiment erscheint jede Person in mindestens 100 aufeinanderfolgenden Bildern).

Auch wenn durch Anpassung der Erkennungsparameter (siehe Abschnitt 3.5) eine Änderung der Fehllerraten möglich ist, so sind doch die Tendenzen der zu wählenden Verfahren und Parameter klar. Aus diesem Grund wird für die markerlose Detektion das SGBM-Verfahren mit einer großen Vergleichsfenstergröße empfohlen. Für das Höhenfeld ist dabei zu beachten, dass es dieses stark glättet und somit untersuchte Formfaktoren einer Person entsprechend ausfallen (siehe Abschnitt 3.5.5).

3.6.6 Trajektorien aus Stereoaufnahmen

Die markerlose Gewinnung von Trajektorien aus Stereoaufnahmen wird in Abschnitt 3.5 beschrieben. Sie dient insbesondere der Laufwegeerfassung in Feldstudien. Der erste Einsatz war jedoch bei der Verbesserung der Trajektorien, die bei den Hermes-Experimenten (Nr. 6) gewonnen wurden. Für die Versuchsreihe Nr. 6 wurden insbesondere aufgrund von Versuchen an Treppen Stereoaufnahmen angefertigt, um die Kameraentfernung bestimmen und 3D-Trajektorien gewinnen zu können. Auch ist bei diesen Versuchen wegen der zum Teil großen Entfernung zur Kamera der Marker sehr einfach aufgebaut gewesen und enthielt keine Information zur Personengröße. Diese konnte ebenfalls mit Hilfe der Stereoaufnahmen in den ebenen Versuchen ermittelt werden (siehe Abschnitt 3.6.2). Zum Testen des markerlosen Extraktionsverfahrens ist im Rahmen dieser Versuchsreihe auch ein Versuch ohne Markierung der Personen durchgeführt worden (siehe Abb. 3.44). Die einzige Feldstudie, also eine Beobachtung im natürlichen Umfeld ohne jegliche Beeinflussung der Personen, wurde bisher am „Tag der Neugier“ am Haupttor des Forschungszentrum Jülich durchgeführt (siehe Abschnitt 3.6.7).

In diesem Abschnitt soll die Qualität der Laufwege der markerlosen Erkennung vornehmlich anhand der Hermes-Experimente (Nr. 6) und virtuellen Personen, insbesondere im Verhältnis zum optimalen Laufweg untersucht werden, da für diese Fälle der exakte Laufweg vorliegt. In Abschnitt 3.5 werden das Erkennungsverfahren und die Kriterien zur Identifikation einer Person für die aus dem Höhenfeld abgeleiteten PES ausführlich besprochen. Die Wahl des Verfahrens und dessen Parameter zur Bestimmung der Disparitätswerte werden in Abschnitt 3.6.5 behandelt. Im Folgenden wird daraufhin das empfohlene SGBM-Verfahren mit einer Fenstergröße von 11 px verwendet.

Als Maß für die Güte der ermittelten Laufwege sollen folgende Punkte untersucht werden:

1. Falsch-Ermittlungs- und Falsch-Negativ-Rate bei der Erkennung der Personen auch in Abhängigkeit der Dichte und des Betrachtungswinkels (siehe Abschnitt 3.6.6.1),
2. Genauigkeit der Positionsbestimmung (siehe Abschnitt 3.6.6.2),
3. Glattheit der Trajektorien (siehe Abschnitt 3.6.6.3).

3.6.6.1 Fehler bei der Erkennung der Personen

Zur Bestimmung der Falsch-Ermittlungs- und Falsch-Negativ-Rate wurden drei reale Versuche bzw. Feldstudien ausgewertet: Experiment an einer 300 cm breiten Ecke (siehe Abb. 3.44) und einer 240 cm breiten T-Kreuzung (siehe Abb. 3.44) aus Versuchsreihe Nr. 6, sowie der Feldstudie im Forschungszentrum Jülich (siehe Abschnitt 3.6.7). Zur systematischen Untersuchung der Abhängigkeit der Fehler von der Personendichte und vom Betrachtungswinkel wurden virtuelle Personen eingesetzt.

Die Ergebnisse sind in Tab. 3.8 zusammengefasst. Es zeigt sich, dass die Fehler der markerlosen Erkennung gering sind, aber die Art des Fehlers nicht für jede Szene identisch ist. Aufgrund von besonderen Objekten und Vorkommnissen überwiegt in der Feldstudie die Falsch-Ermittlungs-Rate z. B. aufgrund mitgeführter Fahrräder und Luftballons. Bei den Experimenten an einer Ecke und T-Kreuzung überwiegt die Falsch-Negativ-Rate, da die Dichte hoch und die Wand und insbesondere die Ecke das Höhenfeld für die Erkennung negativ beeinflussen (siehe Abschnitt 3.5.1).

Das Experiment an der T-Kreuzung, bei denen die Probanden eine Markierung trugen, dient dazu, die Genauigkeit der markerlosen Erkennung im Vergleich zur exakten über den Marker ermittelten Position zu studieren. Zudem kann daran ebenfalls gezeigt werden, dass eine Kombination der Verfahren

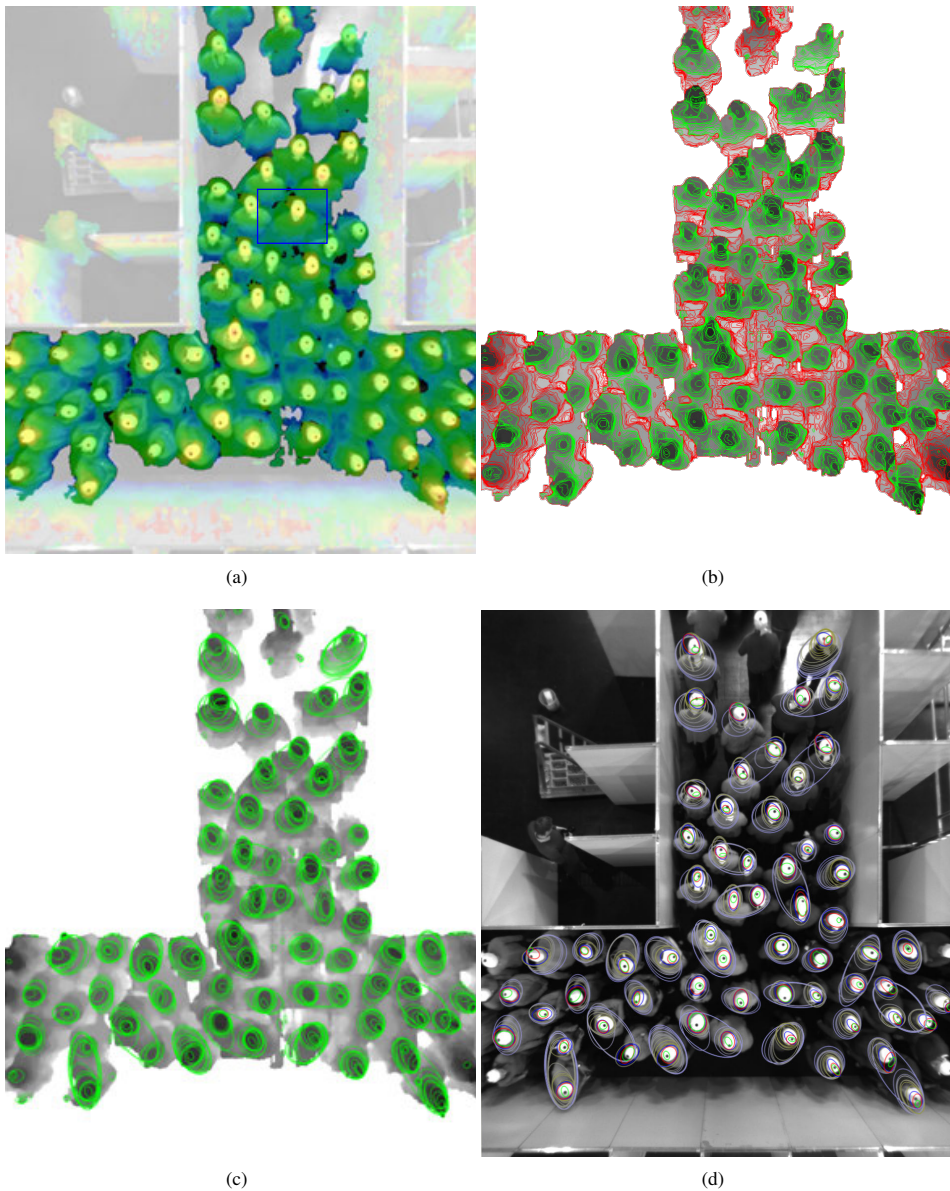


Abb. 3.84: Schritte der markerlosen Personendetektion am Beispiel eines Versuches in einer T-Kreuzung aus Versuchsreihe Nr. 6; (a) Disparität eingeschränkt auf den Bereich des Oberkörpers, der für die Erkennung einer Person herangezogen wird (entspricht der Entfernung zur Kamera von rot: 570 cm – blau: 735 cm); der ausgegraute Bereich entspricht dem bestimmten Hintergrund; (b) Isolinen gleicher Disparität in Abständen von 5 cm im Bereich des Vordergrundes, aufgetragen auf das als Grauwertbild dargestellte Disparitätsbild; grüne Linien umkreisen Bereiche, die näher an der Kamera liegen, rote Linien entsprechend entferntere Bereiche; (c) die grünen Isolinen approximierende Ellipsen; (d) Ellipsen gruppiert nach sich einschließenden Isolinen zur Erstellung der die Personen identifizierende Ellipsenpyramiden; grüne Ellipsen sind innerhalb der Pyramide der Kamera am nächsten, gefolgt von den roten und blauen Ellipsen

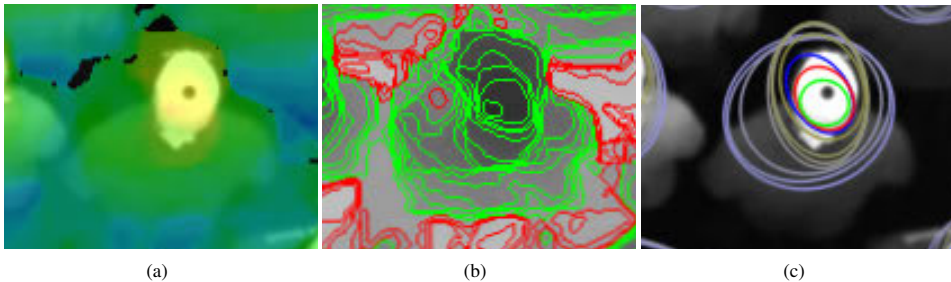


Abb. 3.85: Ausschnitt des blauen Rechtecks in Abb. 3.84 a und folgende Schritte der markerlosen Personendetektion: (a) Disparität, (b) Isolinien, (c) zur Pyramide gruppierte approximierende Ellipsen

die Falsch-Ermittlungs-Rate verringert. Die Dichte beim T-Kreuzungs-Experiment war für die Hermes-Experimente maximal und lag im Kreuzungsbereich und in den zuströmenden Korridoren bei 3-4 Personen/m² (siehe Abb. 4.32). Bei der Anwendung des markerlosen Verfahrens werden die Marker ignoriert. Das ermittelte Höhenfeld unterscheidet sich qualitativ nicht von einem Höhenfeld mit unmarkierten Personen (vgl. Abb. 3.50 und 3.84) und damit auch nicht die darüber bestimmten Isolinien und approximierenden Ellipsen. In Abb. 3.86 sind die mit beiden Verfahren ermittelten Laufwege zu sehen.

Bei den Hermes-Versuchen (Nr. 6) kam es aufgrund des eingesetzten einfach aufgebauten Markers Nr. 2 zu falsch positiven Detektionen, die am Beispiel der T-Kreuzung durch die Kombination beider Verfahren eliminiert werden konnten (siehe Tab. 3.8). Die zu viel erkannten Personen rührten beim Marker Nr. 2 von dem Marker ähnlichen Texturen im Versuchsbereich her und lagen damit für gewöhnlich nicht auf einer Erhebung innerhalb des Höhenfeldes, die als Person identifiziert würde. Bei der kombinierten Nutzung wurden die Marker zur Erkennung herangezogen, jedoch verworfen, wenn sie nicht auf einem durch das markerlose Verfahren detektierten Kopf lagen.

In der Motivation der Einleitung wurde die Notwendigkeit der exakten Personenerkennung in dichten

Versuch, Marker	erkannte Personen	falsch positiv	falsch negativ	Beschleunigung [m/s ²]
Ecke				
markerlos	162	0% (0)	2,4% (4)	3,6 ± 4,8
T-Kreuzung				
markerlos	300	0% (0)	1,0% (3)	6,5 ± 6,8
Marker Nr. 2	304	0,3% (1)	0% (0)	1,2 ± 0,7
Kombination	303	0% (0)	0% (0)	1,2 ± 0,7
Feldstudie				
markerlos	712	2,0% (14)	1,0% (7)	3,4 ± 11,0

Tab. 3.8: Fehler der markerlosen Personenerkennung und der Kombination mit der Erkennung über Marker Nr. 2

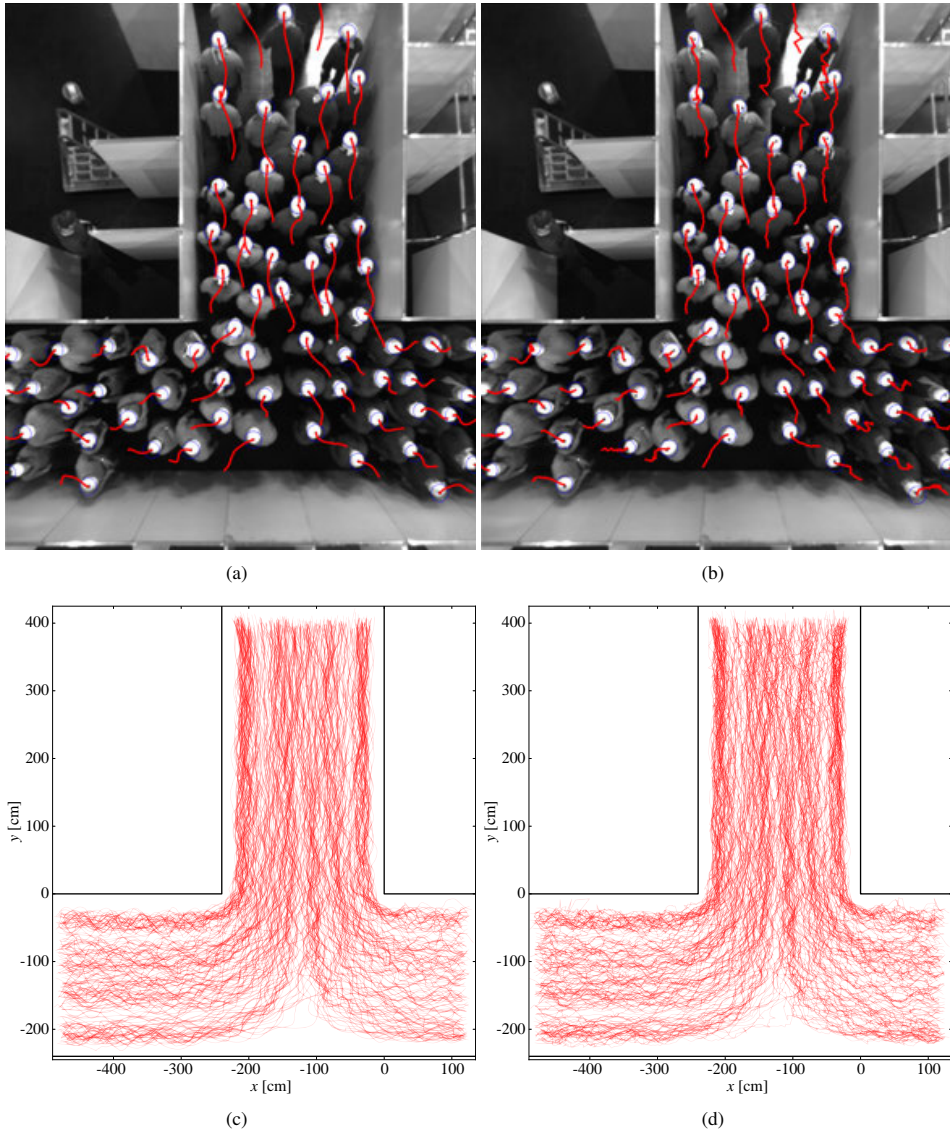


Abb. 3.86: Laufwege von Personen, die eine T-Kreuzung passieren; (a, b) roter Pfad ist Weg der letzten Sekunde; die Mittelpunkte der blauen Kreise beschreiben die aktuell detektierten Positionen; (c, d) alle Trajektorien des Versuchs größenbereinigt projiziert auf die Bewegungsebene; in jedem Frame wurden die Personen (a, c) mithilfe des Hermes-Markers (Nr. 2) oder (b, d) mithilfe des markerlosen Verfahrens auf Basis des Höhenfeldes detektiert; der Laufweg in (a, c) ist glatter bzw. in (b, d) unruhiger aufgrund der Stabilität des auf dem Kopf als Position der Personen herangezogenen Punktes; Abschnitt 4.4.2 zeigt Ergebnisse dieses Versuchs, u. a. sind in Abb. 4.29 die Durchmischung der Ströme sowie die aus zwei Kameraperspektiven zusammengeführten Trajektorien zu sehen

Dichte ρ [1/m ²]	Geschw. v [m/s]	Abb. 3.87	erkannte Personen	falsch positiv	falsch negativ
0,5	1,37	a	34	0% (0)	0% (0)
1,0	1,24	b	77	0% (0)	0% (0)
1,5	1,11	c	104	0% (0)	1,0% (1)
2,0	0,98	d	139	0% (0)	0,7% (1)
2,5	0,85	e	172	0% (0)	0% (0)
3,0	0,72	f	209	0% (0)	0% (0)
3,5	0,59	g	238	0% (0)	0,4% (1)
4,0	0,46	h	251	0% (0)	0% (0)
4,5	0,33	i	235	0% (0)	2,5% (6)
5,0	0,20	j	234	0% (0)	3,7% (9)
5,5	0,20	k	249	0% (0)	6,3% (16)
6,0	0,20	l	259	0% (0)	7,8% (22)

Tab. 3.9: Falsch-Ermittlungs- und Falsch-Negativ-Rate der Personenerkennung in Abhängigkeit von der Personendichte

Menschenmengen beschrieben. Welchen Einfluss die Dichte auf die Fehlerrate hat, ist Tab. 3.9 zu entnehmen.

Hierfür wurden virtuelle Personen mit steigender Dichte in eine Umgebung wie bei den Hermes-Experimenten (Nr. 6) gestellt. Alle Randbedingungen wie u. a. die Kameraparameter entsprachen denen der ebenen Versuche aus Versuchsreihe Nr. 6 (vgl. Abschnitt 3.6.5). Alle Personen besaßen die gleiche Textur, was trotzdem bei der Disparitätsberechnung aufgrund der möglichen Wiederholung keine Fehluordnung hervorrief. Die virtuellen Personengrößen wurden entsprechend den realen Werten mit $176\text{cm} \pm 9\text{cm}$ normalverteilt festgesetzt (vgl. Abschnitt 3.6.2). Entsprechend ihrer Größe wurden die Personen auch in ihrer horizontalen Ausdehnung leicht angepasst. Der Personenabstand wurde entsprechend der angestrebten Dichte eingestellt, wobei die Position in der Bewegungsebene in Abhängigkeit zur Dichte leicht variiert wurde (siehe Abb. 3.87). Dabei konnte es passieren, dass stark unterschiedlich große Personen sehr nah beieinander standen. Die Geschwindigkeit beim Hindurchlaufen unter der Kamera wurde ebenfalls gemäß den maximalen Werten des Fundamentaldiagramms aus Abb. 1.4 festgesetzt. Diese war für alle Personen gleich, da keine Kollisionserkennung implementiert wurde.

Für die Fehlerrate lässt sich sagen, dass keine Person zu viel und bis zu einer Dichte von 4 Personen/m² nur für maximal eine Person kein Laufweg bestimmt werden konnte. Danach steigt die Falsch-Negativ-Rate bei einer Dichte von 6 Personen/m² bis auf 7,8% an. Dies sind sehr gute Werte im Vergleich zu anderen Verfahren (siehe Abschnitt 2.3.11). Ohne starken äußeren Einfluss sind Dichten über 4 Personen/m² selten. Abb. 3.88 zeigt insbesondere bei der maximal untersuchten Dichte, dass nicht in jedem Bild alle Personen gefunden wurden. Da bei der Implementation der Verfahren darauf Wert gelegt wurde, die Anzahl der falsch positiven Detektionen gering zu halten, weil durch das verbindende Tracking nicht jede Person in jedem Frame gefunden werden muss, beziehen sich, wenn nicht anders angegeben, alle Zahlen zu Falsch-Ermittlungs- und Falsch-Negativ-Raten auf Fehldetektionen nach einem kompletten Durchlauf einschließlich des Trackings. Danach liegt für jede Person ihr gesamter Laufweg unter der Kamera vor.

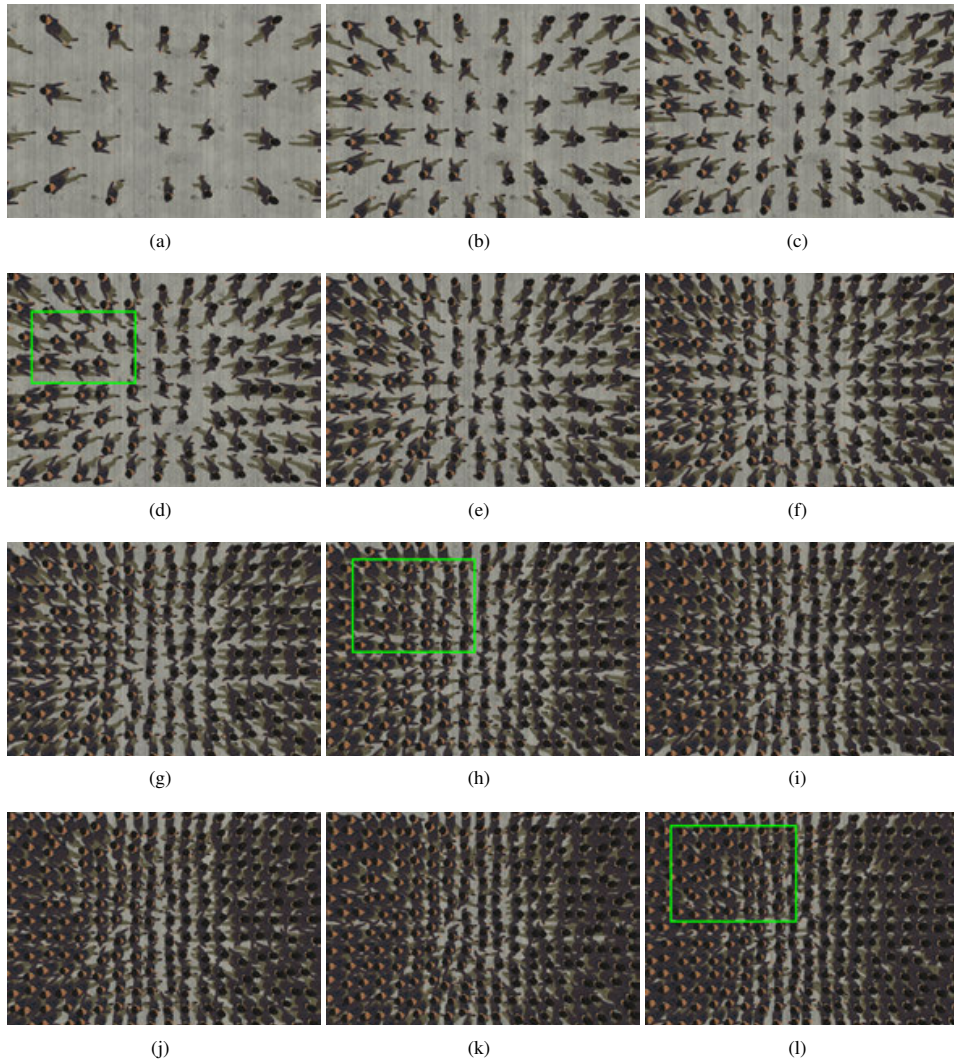


Abb. 3.87: Variierende Personendichte virtueller Personen; die Dichteabstufungen in $0,5/\text{m}^2$ -Schritten sind (a) $0,5/\text{m}^2$, (b) $1,0/\text{m}^2$, (c) $1,5/\text{m}^2$, (d) $2,0/\text{m}^2$, (e) $2,5/\text{m}^2$, (f) $3,0/\text{m}^2$, (g) $3,5/\text{m}^2$, (h) $4,0/\text{m}^2$, (i) $4,5/\text{m}^2$, (j) $5,0/\text{m}^2$, (k) $5,5/\text{m}^2$, (l) $6,0/\text{m}^2$; die Isolines des Kameraabstandes und die Pyramiden der Personen sind für die Bereiche der grünen Rechtecke in Abb. 3.88 zu sehen

Auch wenn für die Herleitung des Verfahrens von einer senkrechten Blickrichtung ausgegangen wird, soll exemplarisch untersucht werden, wie empfindlich die Erkennung auf eine Schrägstellung der Kamera reagiert.

Die Positionsbestimmung nach der Detektion stellt dabei kein Problem dar. Durch die Möglichkeit,

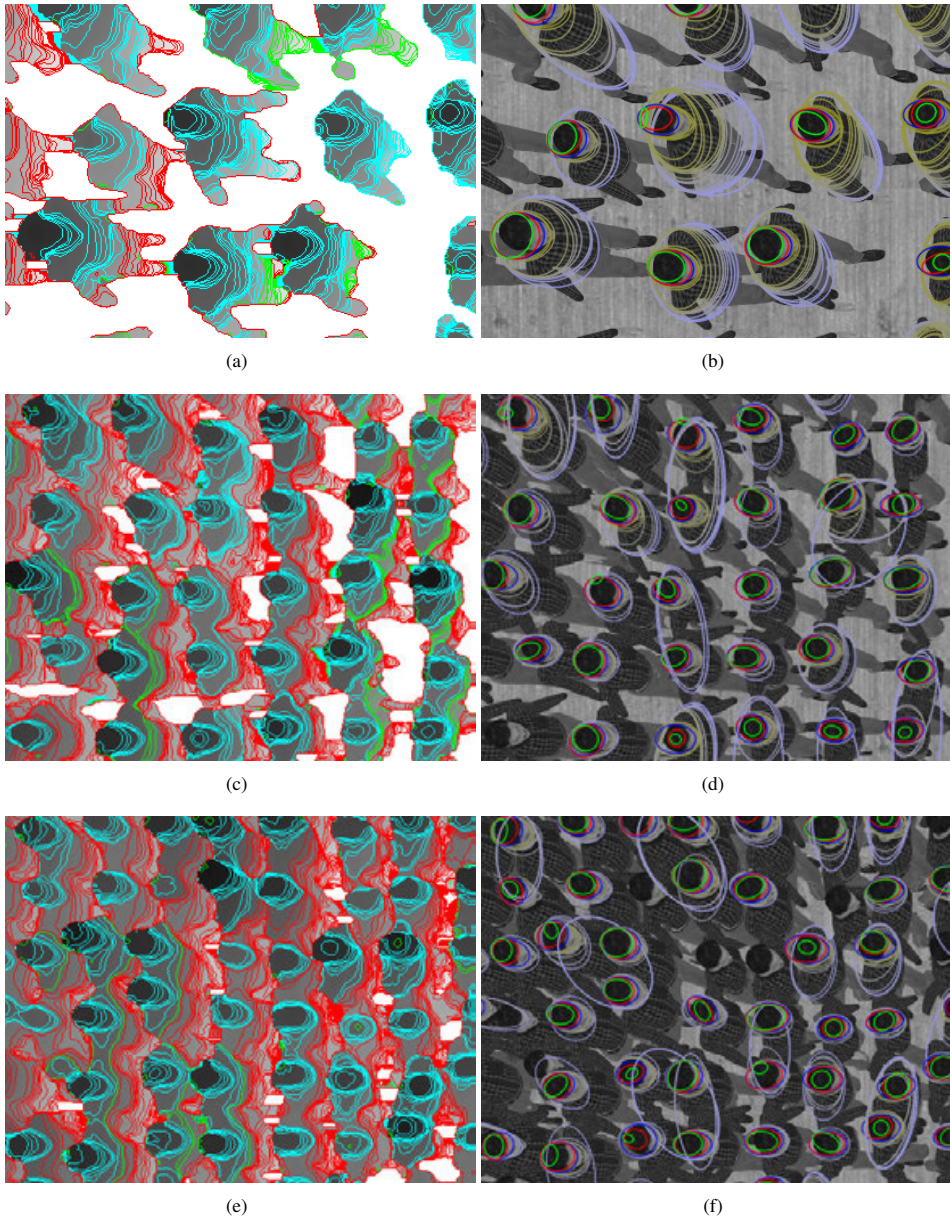


Abb. 3.88: (a,c,e) Isolinen des Kameraabstandes und (b,d,f) Pyramiden virtueller Personen der in Abb. 3.87 grün umrandeten Bereiche mit Personendichten von (a,b) $2,0/\text{m}^2$, (c,d) $4,0/\text{m}^2$ und (e,f) $6,0/\text{m}^2$; die Disparität wurde mit dem SGBM-Verfahren und einer Fenstergröße von 11 px bestimmt; nur die cyanfarbigen Linien werden ggf. zur Konstruktion des PES herangezogen

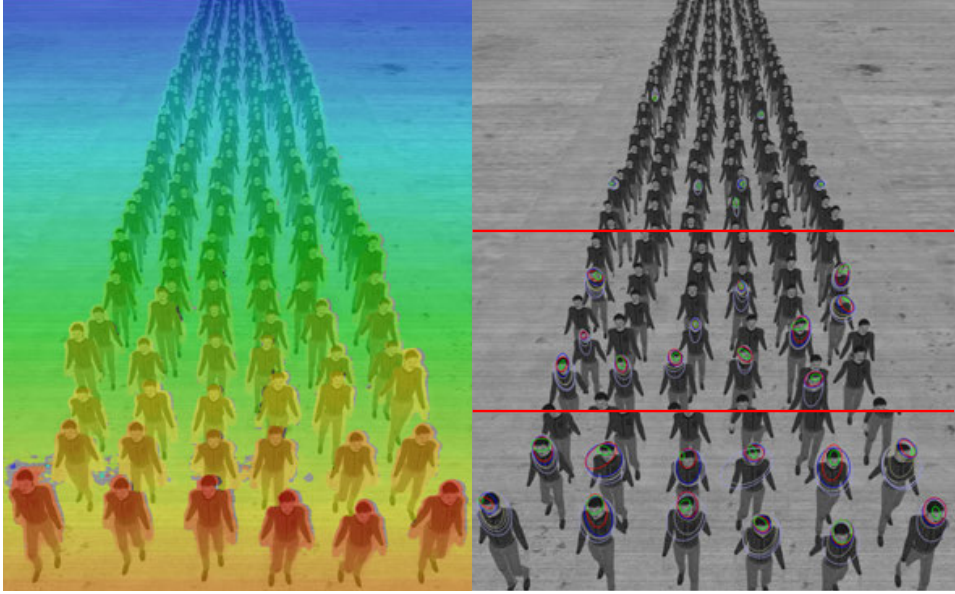


Abb. 3.89: Markerlose Erkennung bei Schrägsicht der Kamera; (links) farbig überlagerte Disparität von 5 px (blau) bis 38 px (rot); (rechts) drei Bildbereiche, für die getrennt eine Erkennung durchgeführt wurde, und mittels PES markierte Detektionen

aus der Disparität und den Kameraparametern die 3D-Koordinate eines Bildpunktes zu bestimmen, ist man beim markerlosen Verfahren nicht auf eine exakt senkrechte Aufhängung und ein manuell eingestelltes Koordinatensystem auf dem Boden (siehe Abschnitt 3.1.1) angewiesen. In Kamerakordinaten ergibt sich für die Pixelkoordinate $(u \ v)^T$ unter Vernachlässigung der Rotation und Translation der Kamera zur Einbettung in das Weltkoordinatensystem gemäß Gleichung 3.24 und 3.18:

$$\bar{x}' = \frac{sb}{\delta(u)} \begin{pmatrix} \frac{1}{s_x}(u - c_x) \\ \frac{1}{s_y}(v - c_y) \\ 1 \end{pmatrix}. \quad (3.37)$$

Diese Möglichkeit der direkten Positionsbestimmung ist ebenfalls in *PeTrack* implementiert. Die Bestimmung der Schrägstellung der optischen Achse kann bei ebenen Versuchen unter der Annahme erfolgen, dass jede Trajektorie parallel zur Bewegungsebene verläuft (zur Genauigkeit vgl. Abschnitt 3.6.2). Bei Versuchen an Treppen wurde diese Annahme auf einen ebenen Bereich in der Aufnahme eingeschränkt (z. B. Podest an der Außentreppe (siehe Abb. 4.3 c und 4.9) oder Ebene direkt vor dem Ausgang des Mundlochs (siehe Abb. 3.70 folgende und 4.8)) getroffen (vgl. Abschnitt 3.3.5).

Die Abhängigkeit der Identifikation von der Schrägstellung soll wieder anhand virtueller Personen durchgeführt werden. Abb. 3.89 zeigt das untersuchte Szenario. Von allen sonstigen virtuellen Szenen weicht diese nur dadurch ab, dass die optische Achse nicht parallel zum Lot, sondern in einem Winkel von 57° zum Lot ausgerichtet ist und damit eher der Ausrichtung einer Überwachungskamera gleicht. Für die virtuelle Kamera mit einem vertikalen Öffnungswinkel von 49° entsprechend der realen Stereokamera liegt die untere Bildkante bei einem Blickwinkel von $32,5^\circ$ und damit genau beim

maximalen horizontalen Öffnungswinkel der senkrechten Sicht. Trotzdem unterscheidet sich das Höhenprofil an der unteren Bildkante vom Profil einer Kamera, die senkrecht ausgerichtet ist, da die Disparität die Entfernung parallel zur optischen Achse wiedergibt. Wäre das Höhenprofil unabhängig von der Rotation der Kamera, müsste z. B. die Disparität des Bodens auch für die Schrägsicht wie bei der senkrechten Ausrichtung gleich sein.

Bei einer Personendichte von $1/\text{m}^2$ bewegten sich die Personen auf die Kamera zu. Für die Fehlerbeurteilung wurde die Szene in drei Bildbereiche aufgeteilt, für die getrennte Erkennungsläufe durchgeführt wurden. Im untersten Bildbereich von Abb. 3.89 wurden bei 63 erkannten Personen 1 (1,6%) zu viel und 4 (6,3%) zu wenig erkannt. Im mittleren Bildbereich lag die Fehlerrate bei 11 (15,9%) zu viel erkannten Personen und 1 (1,4%) zu wenig detektierten Person bei insgesamt 69 Identifikationen. Für den oberen Bildbereich kann keine sinnvolle Fehlerrate angegeben werden, da die Personen nur selten erkannt wurden, so dass durch die Bereinigung während des Trackings die ermittelten Laufwege verworfen werden, da in weniger als 20% der Einzelbilder die Personen identifiziert wurden (siehe Abschnitt 3.3.4).

3.6.6.2 Genauigkeit der Position

Für den Punkt 2 der Liste der Qualitätsanalyse von Seite 148, die die Genauigkeit der Positionsbestimmung beschreiben soll, wird die Position, die mittels des markerlosen Verfahrens ermittelt wurde, mit der Position verglichen, die über einen strukturierten Marker ermittelt wurde. Ist das Strukturelement von jedem Blickwinkel der Kamera aus komplett einzusehen und eben, so beschreibt der Mittelpunkt der approximierenden Ellipse immer denselben Punkt auf dem Kopf der Person. Der feste Punkt kann jedoch aufgrund der Variation beim Tragen der Kopfbedeckung und auch durch Bewegung des Kopfes unterschiedliche Positionen in der Bewegungsebene relativ zur Position der Person beschreiben. Beide Verschiebungen sind jedoch für den gesamten Beobachtungsbereich gleich bzw. verändern sich unabhängig vom Blickwinkel.

Die bereits mehrfach verwendete virtuelle Person verändert die Kopfhaltung nicht. In Abb. 3.90 ist zu sehen, wie dieser Person durch Änderung der Textur der Hermes-Marker (Nr. 2) aufgesetzt wurde. Zudem zeigt die Abbildung die Differenz in der ermittelten Position der Person mit Hilfe der Verfahren ohne und mit Marker. Die Differenz für den gesamten Laufweg ist Abb. 3.79 zu entnehmen. Da die Kopfbedeckung nicht kopfmittig aufgebracht wurde, sondern zum Gesicht orientiert war, wurde für den Vergleich die Trajektorie, die mittels Hermes-Marker ermittelt wurde, global um $(3,6\text{ cm } 0,5\text{ cm})^T$ verschoben. Dies wäre nicht nötig gewesen, wenn der Hermes-Marker exakt in der Mitte des Kopfes aufgebracht worden wäre. Aus diesem Grund ist in der linken Spalte von Abb. 3.90 kein Fehler zu beobachten. Durch die Verschiebung würde dieser ansteigen und dafür sich der Fehler der rechten Spalte verringern. Der Fehler verhält sich sehr ähnlich dem Fehler des Farbmarkers, der sich ebenfalls von der Bildmitte ab bei steigendem Blickwinkel erhöht bzw. die ermittelte Position zur Bildmitte wandert (vgl. Abb. 3.79 und 3.91). Die Abbildungen sind identisch skaliert, so dass beim Vergleich auffällt, dass der Fehler beim markerlosen Verfahren größer ist und damit auch der Mittelwert der mikroskopischen Geschwindigkeit gemäß Gleichung 3.35 kleiner ausfällt.

Die Verschiebung zur Bildmitte entsteht, da das perspektivische Höhenfeld genutzt und, wie in Abb. 3.80 zu sehen, dadurch die Isolinien einer Höhe entlang von nicht einsehbaren Bereichen laufen. Diese Sprünge, an denen alle Isolinien verlaufen, deren benachbarte Pixel den Isowert einschließen, sind z. B. an der gesamten Grenze einer Person zum Hintergrund vorzufinden, so auch an der Stirn oder dem Hinterkopf, von dem man je nach Perspektive unterschiedliche Bereiche sieht.

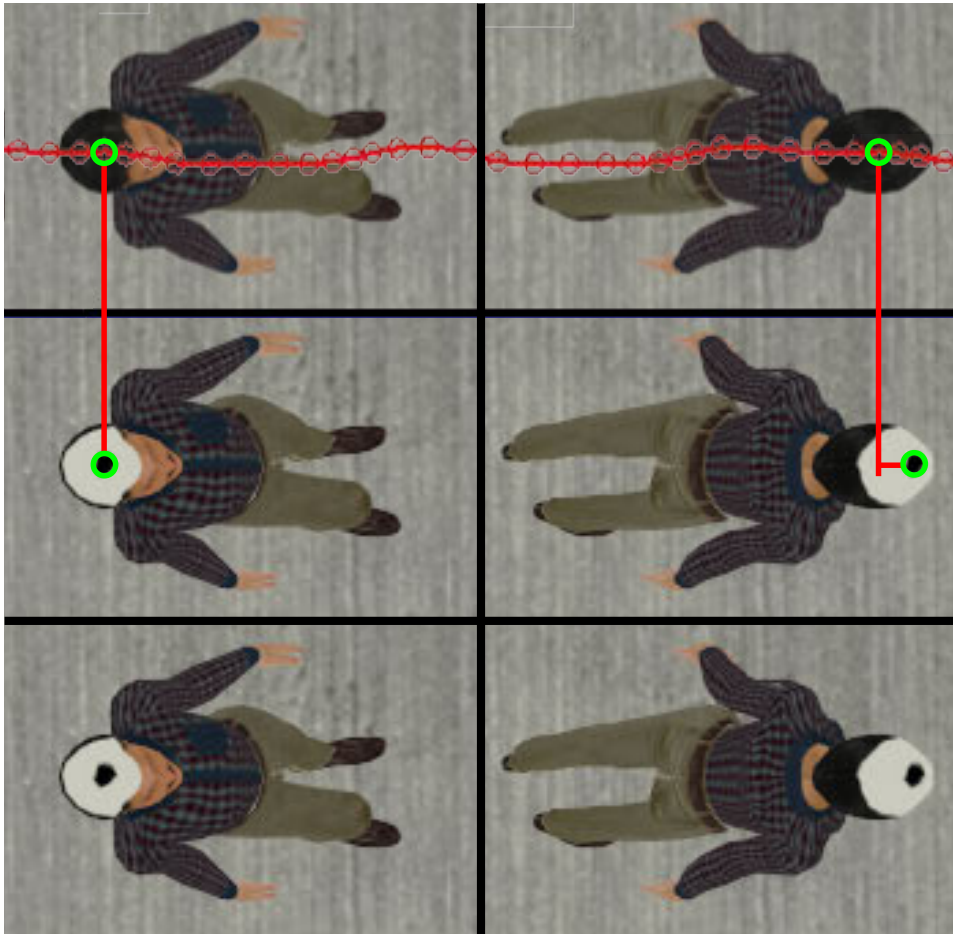


Abb. 3.90: Differenz der ermittelten Position zwischen dem markerlosen Verfahren und der Erkennung mittels des Hermes-Marker (Nr. 2); der Betrachtungswinkel kann der zweiten (-24°) und letzten Spalte ($28,2^\circ$) in Abb. 3.80 entnommen werden, bei der die virtuelle Person komplett zu sehen ist; der linke Ausschnitt entspricht Abb. 3.81; die Mittelpunkte der grünen Kreise zeigen den Detektionsort (oben) des markerlosen Verfahrens und (mitte) den, der durch den Hermes-Marker mit einem schwarzen Punkt als strukturierendes Element ermittelt wurde; (unten) Platzierung des Hermes-Markers auf dem Kopf der virtuellen Person über Anpassung der Textur

Wie beim Farbmarker kann die zur Bildmitte verschobene Position durch eine Skalierung so korrigiert werden, dass sie der optimalen sehr nahe kommt. Da das über die Disparität ermittelte Höhenfeld, wie in Abb. 3.82 oder 3.51 folgende zu sehen, u. a. aufgrund von texturabhängigen Schwankungen in der Disparität nicht analytisch fassbar ist, wurde ein Faktor heuristisch ermittelt. Sei \vec{c}' die metrische Bildmitte, so würde der Faktor

$$1,02 - 0,006 \arctan \frac{\|p_i(t) - \vec{c}'\|}{a - \bar{p}^h} \quad (3.38)$$

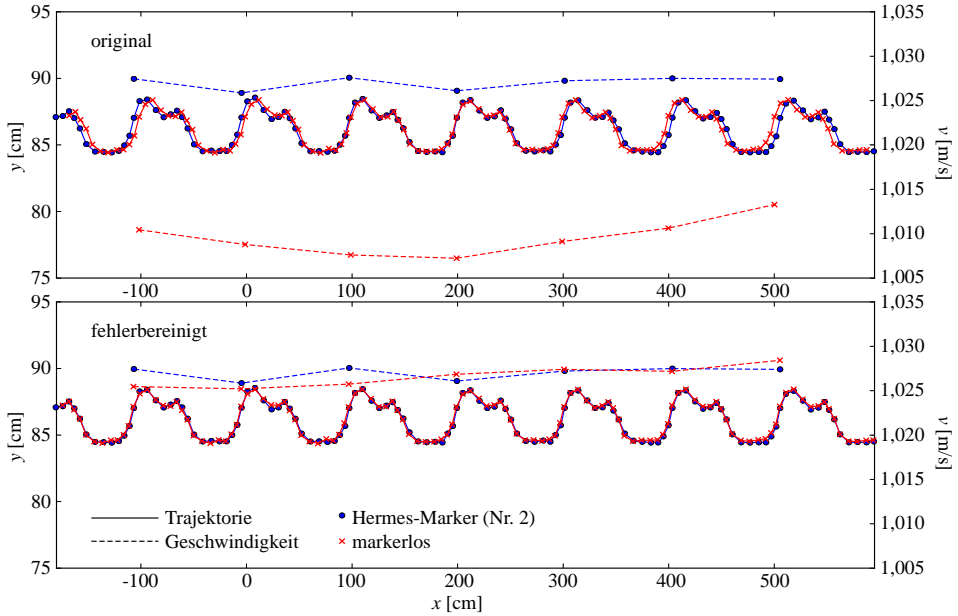


Abb. 3.91: Trajektorie und über eine Doppelschrittlänge gemittelte Geschwindigkeit derselben virtuellen Person, die über den Hermes-Marker (Nr. 2) und das Höhenprofil erkannt wurde; (oben) sind die bis auf eine Gesamtverschiebung von $(3,6\text{ cm } 0,5\text{ cm})^T$ erhaltenen Originalpositionen und (unten) der für das markerlose Verfahren mit dem Faktor aus Gleichung 3.38 angepasste Laufweg und die resultierende Geschwindigkeit entsprechend Gleichung 3.35 abgebildet

in Abhängigkeit vom Abstand zur Bildmitte $\|p_i(t) - \vec{c}'\|$ den optimierten Laufweg im unteren Bereich der Abb. 3.91 erzeugen und damit auch die mikroskopische Geschwindigkeit auf die reale Geschwindigkeit anheben.

Dieser Fehler bzw. Abstand ist auch im realen Experiment der T-Kreuzung zu beobachten. Sei $p_i^*(t)$ der mit Hilfe des Markers 2 ermittelte Laufweg der Person $i \in \mathbb{N}_N$ für die Zeit $t \in [{}^*t_i^{\text{in}}, {}^*t_i^{\text{out}}]$ und $p_i(t)$ der ohne Marker extrahierte Laufweg der gleichen Person für die Zeit $t \in [t_i^{\text{in}}, t_i^{\text{out}}]$. Dann ist $[t_i^{\text{in}}, t_i^{\text{out}}]$ mit $t_i^{\text{in}} = \max({}^*t_i^{\text{in}}, t_i^{\text{in}})$ und $t_i^{\text{out}} = \min({}^*t_i^{\text{out}}, t_i^{\text{out}})$ der Zeitraum, in dem Person i mit beiden Verfahren erfasst wurde. Der gerichtete Abstand über die Zeit ist damit definiert über

$$d_i(t) = p_i^*(t) - p_i(t) \quad \forall i \in \mathbb{N}_N \wedge t \in [t_i^{\text{in}}, t_i^{\text{out}}] \quad (3.39)$$

und der durchschnittliche Abstand aller N Personen bei einem Einzelbildabstand von Δt über

$$\bar{d} = \frac{1}{N} \sum_{i=1}^N \frac{\Delta t}{t_i^{\text{out}} - t_i^{\text{in}} + \Delta t} \sum_{t=t_i^{\text{in}}}^{t_i^{\text{out}}} \|d_i(t)\|. \quad (3.40)$$

Für den untersuchten Lauf des Experimentes an der T-Kreuzung ergibt sich ein durchschnittlicher Abstand von

$$\bar{d} = 3,8\text{ cm} \pm 1,0\text{ cm}.$$

Um die örtliche Verteilung des Abstandes zu veranschaulichen, soll dieser auf ein regelmäßiges Gitter aufgeteilt werden. Der durchschnittliche gerichtete Abstand innerhalb einer Gitterzelle mit den

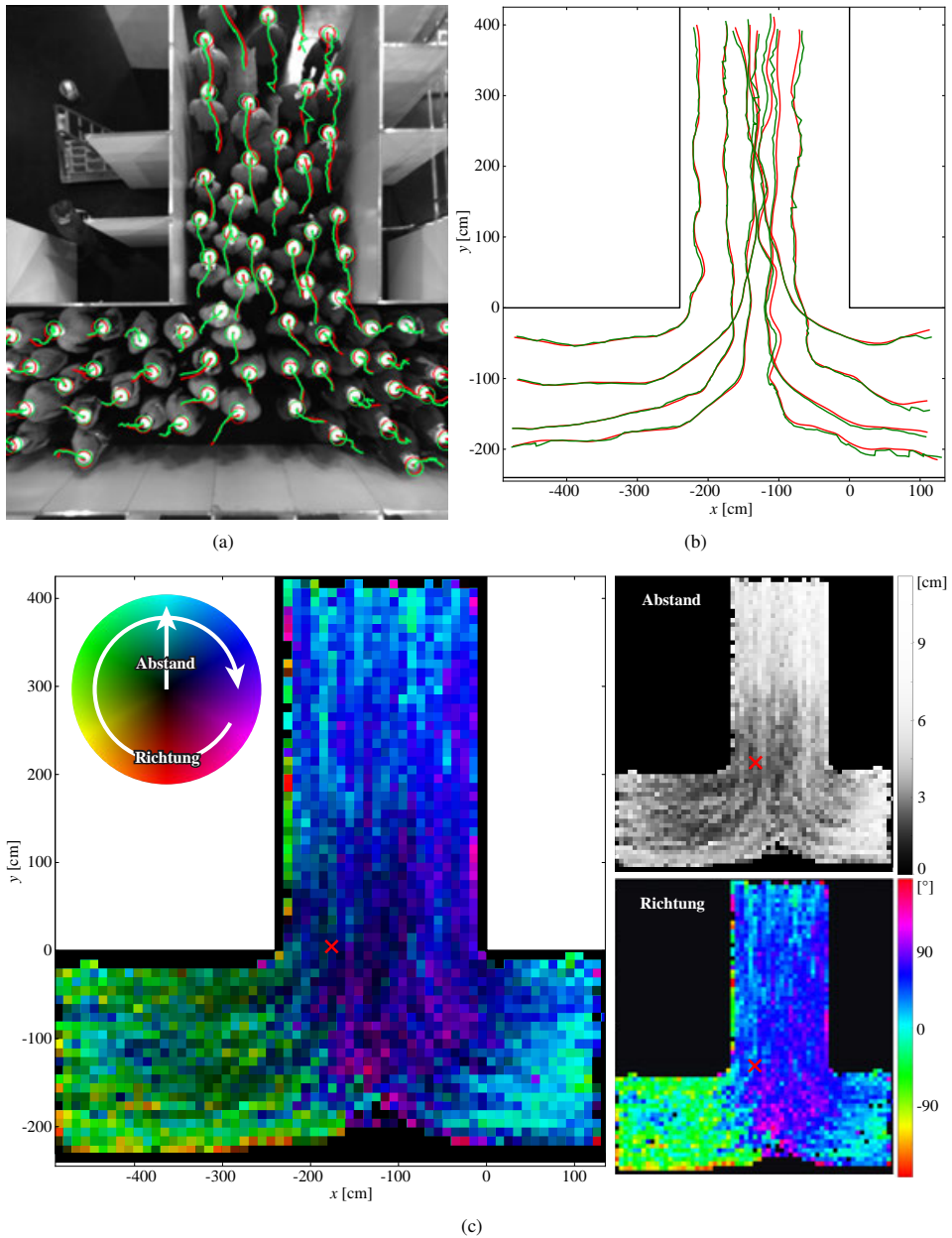


Abb. 3.92: Vergleich der Laufwege, die mit (rot) und ohne Marker (grün) ermittelt wurden; (a) die Trajektorien beschreiben die Positionen in der letzten Sekunde; (b) Auswahl von kompletten Trajektorien größenbereinigt projiziert auf die Bewegungsebene; (c) der Betrag aller gerichteten Abstände zwischen entsprechenden Trajektorien als Helligkeit gemäß Gleichung 3.41 zusammengefasst auf einem Raster mit einer Zellengröße von $10\text{ cm} \times 10\text{ cm}$; der Farbton kodiert die Richtung der Verschiebung; das rote Kreuz markiert die Bildmitte der ursprünglichen Aufnahme

Außenmaßen $G = \{(x, y)^T \in \mathbb{R}^2 \mid x \in [x_{\min}, x_{\max}] \wedge y \in [y_{\min}, y_{\max}]\}$ ist gegeben durch

$$d(G) = \sum_{i=1}^N \frac{1}{|d_i(G)|} \sum_{d_i(t) \in d_i(G)} d_i(t). \quad (3.41)$$

Dabei beschreibt $d_i(G) = \{d_i(t) \mid t \in [t_i^{\text{in}}, t_i^{\text{out}}] \wedge p_i^*(t) \in G\}$ die Menge an Abständen, die innerhalb der Gitterzelle G liegen, und $|d_i(G)|$ die Mächtigkeit dieser Menge.

Abb. 3.92 c zeigt die Verteilung des gerichteten Abstandes bei einem Versuch in der T-Kreuzung unterteilt auf ein Raster mit einer Zellengröße von $10\text{cm} \times 10\text{cm}$. Der Farbton zeigt den Richtungswinkel des Abstandsvektors $d(G)$ und die Helligkeit den Betrag des Abstandes $\|d(G)\|$ an. Das große Bild enthält über den Hue- und Value-Wert der Farbe im HSV-Farbraum beide vereinten Informationen. Schwarze Zellen beinhalten keine Werte, da kein Laufweg über diese Bereiche verläuft. Da in Abb. 3.92 c nur ein Bildausschnitt dargestellt wird, wurde mit dem roten Kreuz die Bildmitte markiert. Die den Betrag kodierende Helligkeit zeigt deutlich, dass auch beim realen Experiment der Abstand bzw. Fehler zum Bildrand hin zunimmt. In der Bildmitte ist die Sättigung am geringsten, also der Abstand zwischen den Trajektorien am kleinsten. Es ist eine klare Abhängigkeit zwischen dem Ort des Fehlers und der Richtung der Verschiebung zu erkennen. Die die Richtung des Abstandes kodierende Farbe zeigt wie bei der virtuellen Person bis auf den rechten unteren Abschnitt des Bildes von der Bildmitte in Richtung des Ortes der Person. Starke Schwankungen in Betrag und Richtung der Abstände nah der Wände rühren von der Tatsache her, dass auch Zellen mit nur einer Person visualisiert wurden und damit einer großen Streuung unterworfen waren.

In Abb. 3.92 a ist der Laufweg der letzten Sekunde aller abgebildeten Personen mit beiden Verfahren zum Vergleich abgebildet. Für acht ausgewählte Personen sind in Abb. 3.92 b die kompletten Trajektorien, die mit (rot) und ohne Marker (grün) ermittelt wurden, größenbereinigt auf die Bewegungsebene projiziert. Dabei ist der wachsende Fehler zum Rand ebenfalls zu beobachten. Zudem ist in a und c der Unterschied in der Glattheit der Trajektorien zu erkennen, die im folgenden Abschnitt untersucht wird.

Bei Betrachtung der Geschwindigkeit ist in Abb. 3.93 für das T-Kreuzungs-Experiment wie für die virtuelle Szene ein Anstieg der Geschwindigkeit von der Bildmitte zum Bildrand zu beobachten. Jeder Datenpunkt wurde gemäß Gleichung 3.35 gemittelt. Die Ausgleichsgrade für die Geschwindigkeit in Abhängigkeit zum Abstand zur Bildmitte \bar{c}' ist für das als Grundwahrheit herangezogene Verfahren unter Nutzung des Hermes-Markers (Nr. 2)

$$0,54 \frac{\text{m}}{\text{s}} + 0,014 \frac{\|p_i^*(t) - \bar{c}'\|}{\text{s}}$$

und für das markerlose Verfahren

$$0,54 \frac{\text{m}}{\text{s}} + 0,04 \frac{\|p_i(t) - \bar{c}'\|}{\text{s}}.$$

Die identische Geschwindigkeit in der Bildmitte rührt von der zufälligen Aufhebung der verfahrensbedingten geringeren Geschwindigkeit und der die Geschwindigkeit erhöhenden unruhigeren Trajektorien her. Die Mittelung über eine Sekunde erlaubt es in der Realität nicht, jede Person wieder auf demselben Fuß zu erwischen. Zwar kann eine angepasste Gleichung 3.38 mit verschwindendem konstanten Faktor

$$-0,14 \arctan \frac{\|p_i(t) - \bar{c}'\|}{a - \bar{p}^h}$$

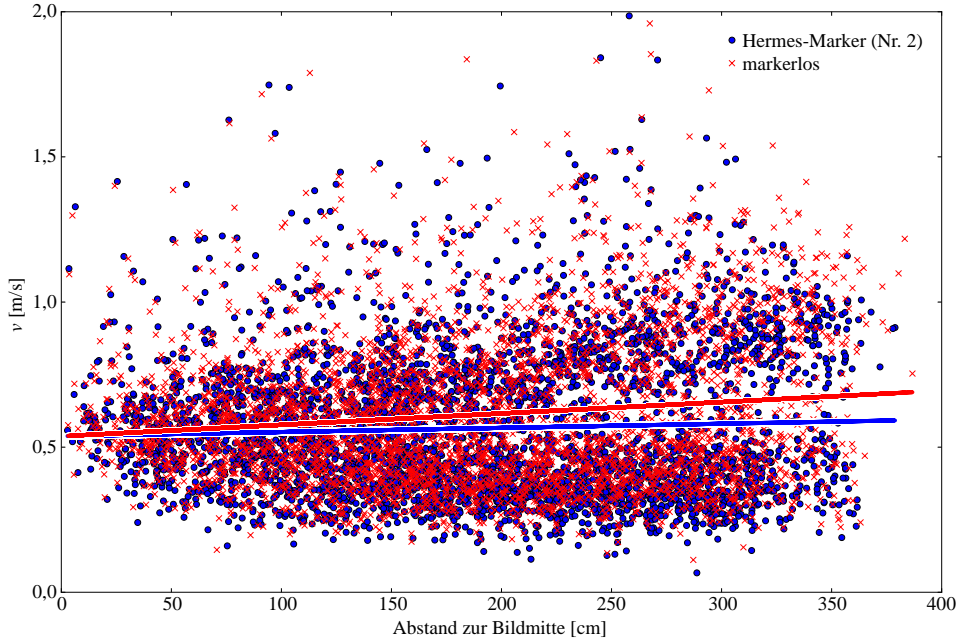


Abb. 3.93: Geschwindigkeit in Abhängigkeit vom Abstand zur Bildmitte ermittelt unter Verwendung von Gleichung 3.35 für Laufwege extrahiert über Verfahren mit (blau) und ohne Marker (rot); die rote Ausgleichsgerade des markerlosen Verfahrens steigt zum Bildrand stärker an als die blaue Gerade des Verfahrens, das den Hermes-Marker (Nr. 2) nutzt (vgl. Abb. 4.32)

die Ausgleichsgeraden in Deckung bringen, jedoch nicht die eigentliche Position verbessern, da die Skalierung nur bewirkt, dass die Position zur Bildmitte zusammengezogen wird, um die Geschwindigkeit zu verringern und damit in Übereinstimmung zu bringen.

Die geringere Anzahl an Datenpunkten in Abb. 3.93 wie auch PES in Abschnitt 3.5.5 in der Bildmitte liegt an der konzentrisch ansteigenden Personenzahl bei der Betrachtung von abstands- oder winkelabhängigen Verteilungen in ebenen Versuchen. Für äquidistant große Intervalle von Abständen zur Bildmitte von Δx oder Winkel zur optischen Achse von $\Delta \alpha$ ergibt sich ein faktorieller Anstieg des Flächeninhalts respektive der Personenzahl für das $(i+1)$ -te Intervall von

$$\frac{\pi((i+1)\Delta x)^2 - \pi(i\Delta x)^2}{\pi(i\Delta x)^2 - \pi((i-1)\Delta x)^2} = \frac{2i+1}{2i-1} \quad (3.42)$$

bzw.

$$\frac{\pi(a \tan((i+1)\Delta \alpha))^2 - \pi(a \tan(i\Delta \alpha))^2}{\pi(a \tan(i\Delta \alpha))^2 - \pi(a \tan((i-1)\Delta \alpha))^2} = \frac{(\tan((i+1)\Delta \alpha))^2 - (\tan(i\Delta \alpha))^2}{(\tan(i\Delta \alpha))^2 - (\tan((i-1)\Delta \alpha))^2} \quad (3.43)$$

Der abstandsabhängige Faktor verringert sich bei steigendem i kontinuierlich, der winkelabhängige fällt bis 30° und wächst danach an. Dass die Menge an Datenpunkten in Abb. 3.93 nicht entsprechend Gleichung 3.42 weiter ansteigt, liegt an der Eingrenzung der Bewegung auf das Innere des Versuchsaufbaus.

Die Geschwindigkeit weist zum Bildrand zwei Zustände auf. Diese spiegeln die zu- und abführenden

Korridore der T-Kreuzung wieder, deren unterschiedlichen Geschwindigkeitsprofile auch in Abb. 4.32 zu beobachten sind. Die ansteigende Geschwindigkeit eines dünneren Zweiges entspricht dem abfließenden Strom aus der T-Kreuzung. Das Intervall, in dem der Großteil der Geschwindigkeiten liegt, entspricht mit $[0, 2, 1]$ dem in Abb. 4.32.

3.6.6.3 Glattheit der Trajektorien

Um eine Aussage über die Glattheit einer Trajektorie machen zu können und damit die Stabilität des Detektionsortes auf dem Kopf zu quantifizieren, soll der Betrag der mikroskopischen Beschleunigung herangezogen werden. Dieser fällt umso kleiner aus, desto ruhiger die bestimmte Position auf der gleichen Stelle des Kopfes verbleibt. Eine kleine Positionsänderung auf dem Kopf verursacht auch eine geringe Beschleunigung.

Der Betrag der mikroskopischen Beschleunigung der Person i zum Zeitpunkt t_j entlang ihres Laufweges p_i wird über aufeinanderfolgende Frames einer Bildsequenz berechnet:

$$a_i(t_j) = \frac{\|(p_i(t_{j+1}) - p_i(t_j)) - (p_i(t_j) - p_i(t_{j-1}))\|}{(\Delta t)^2}. \quad (3.44)$$

Die Beschleunigung wird mikroskopisch genannt, um auszudrücken, dass sie entlang des detektierten Laufweges und nicht entlang der Hauptbewegungsrichtung bestimmt wird (vgl. Abschnitt 4.3).

Auch die eigentliche Bewegung verursacht beim gleichmäßigen Gehen eine Beschleunigung, zum einen niederfrequente Schwankungen durch die Verlagerung von einem Bein auf das andere Bein (siehe Abb. 3.68) und zum anderen höherfrequente Änderungen der Beschleunigung innerhalb eines Schrittes, die ebenfalls auf den gesamten Bewegungsapparat zurückzuführen sind. Diese Grundbeschleunigung lässt sich über die Trajektorien gewinnen, die mittels Strukturelementen auf dem Kopf immer die gleiche Position verfolgen. Die Bewegung des Kopfes einiger Probanden während des Gehens fließt dabei mit in die Beschleunigung ein. Dieses geschieht für alle Verfahren gleichermaßen.

Abb. 3.92 a zeigt im direkten Vergleich den Unterschied im Verlauf der Trajektorien zwischen der Detektion mit und ohne Marker (vgl. Abb. 3.86). Damit der Laufweg etwas über die Stabilität des detektierten Ortes aussagt, wurde für die Laufwege in jedem Frame eine Detektion der Personen vorgenommen und nicht über mehrere Frames hinweg nur verfolgt, was dazu führen würde, dass auch das markerlose Verfahren einen glatteren Laufweg erzeugen würde. Die auch in Abb. 3.92 b insbesondere zum Rand zu beobachtenden teilweise starken Schwankungen führen für das untersuchte T-Kreuzungs-Experiment zu einem Betrag der mikroskopischen Beschleunigung von $6,5 \text{ m/s}^2 \pm 6,8 \text{ m/s}^2$ für das markerlose Verfahren. Der Vergleichswert liegt bei $1,2 \text{ m/s}^2 \pm 0,7 \text{ m/s}^2$ bei Nutzung des Hermes-Markers, der damit die untere Schranke der Beschleunigung für diesen Versuch darstellt.

Das markerlose Verfahren führt bei einem Experiment an einer 3 m breiten Ecke aus Versuchsreihe Nr. 6 (siehe Abb. 3.44) zu einer Beschleunigung von $3,6 \text{ m/s}^2 \pm 4,8 \text{ m/s}^2$. Die geringere Unruhe im Laufweg und damit größerer Stabilität der detektierten Position könnte an der gleichmäßigen Bewegung der Probanden liegen, da sie ungehindert den Versuchsaufbau ohne Eingliederung in einen anderen Personenstrom durchqueren konnten.

Für die Feldstudie ist ein Betrag der mikroskopischen Beschleunigung von $3,4 \text{ m/s}^2 \pm 11,0 \text{ m/s}^2$ gemessen worden. Sie besitzt damit im Schnitt die gleiche durchschnittliche Beschleunigung wie das Eck-Experiment. Die große Standardabweichung rührt insbesondere von der Überbelichtung her, die viele Oberkörper zwar noch erkennen ließ, aber zu einer großen Schwankung in der ermittelten Position führte.

Zur Verringerung der verfahrensbedingten Unruhe in den Trajektorien könnte die Bestimmung der Position über die Achse der Mittelpunkte des PES beitragen (siehe Abschnitt 3.5.5).

Auch könnten die in Abschnitt 4.3 beschriebenen, den Laufweg glättenden Methoden (Wendepunktinterpolation, Frequenzanalyse, Kopfausrichtung, Größenänderung) zu einer Verbesserung der Ergebnisse führen und neben der Bestimmung der Hauptbewegungsrichtung auch die niederfrequenten Schwankungen beseitigen helfen.

3.6.7 Feldstudie

Das markerlose Verfahren wurde während der Hermes-Experimente (Nr. 6) bei der Detektion mit Markern nur zur Verringerung der Falsch-Ermittlungs-Rate verwandt (siehe Abschnitt 3.6.6.1). Bei einer Feldstudie wurde es das erste Mal beim „Tag der Neugier“ des Forschungszentrum Jülich am 29.09.2013 im Vergleich mit anderen Produkten und Verfahren eingesetzt.

Bei der Studie sollte am Haupteingang des Forschungszentrums die ein- und austretende Personenzahl erfasst werden. Über den Tag diente diese Zahl der Überwachung der Anzahl der auf dem Gelände befindlichen Besuchern. Aufgrund der tragischen Ereignisse auf der Loveparade in Duisburg sollte diese jederzeit bekannt sein und als Kriterium für eine Schließung des Geländes herangezogen werden. Nach der Veranstaltung sollte damit der Presse eine verlässliche Besucherzahl genannt werden können, die bei vorherigen Veranstaltungen dieser Art nur grob geschätzt wurde.

Der Sicherheitsdienst des Forschungszentrums hat hierfür das Produkt TrueView People Counter [513] der Firma Cognimatics beschafft. Der Hersteller hat bezüglich der Zuverlässigkeit der Anlage eine Erkennungsrate von über 90% angegeben. Aufgrund meiner Erfahrungen u. a. im Hermes-Projekt mit zu hoch angegebenen Erkennungsraten der Hersteller wurden zum Vergleich zwei weitere Produkte bzw. Verfahren eingesetzt. Zum einen war die Firma nvl² mit ihrem Produkt MoPeCo [515] bereit, sich einem Vergleich zu stellen, und zum anderen konnte das in dieser Arbeit vorgestellte markerlose Verfahren das erste Mal in einer realen Umgebung getestet werden.

Die Zählkamera des Produktes TrueView People Counter wurde vom Sicherheitsdienst gemäß Bedienungsanleitung und der Laserscanner des Produktes MoPeCo von der Firma nvl² aufgebaut und eingerichtet. Zur Aufhängung der Erfassungssysteme wurde ein Traversensor mit einer Durchgangsbreite von 6,42 m aufgebaut, durch das alle Besucher hindurchgehen mussten (siehe Abb. 3.96 und 3.94 a). Neben den drei Erfassungssystemen wurde eine HDV-Kamera (Nr. 7) mit Weitwinkelkonverter für die manuelle Überprüfung installiert (Sichtbereich siehe Abb. 3.94 b). Die HDV-Kamera musste hochkant und schräg ausgerichtet werden, um den gesamten Durchgang einsehen zu können. Alle anderen Systeme besaßen eine senkrechte Blickrichtung. Die Stereokamera war dadurch nur in der Lage, die Hälfte des Durchgangs zu erfassen. Alle Systeme (siehe Abb. 3.95) lieferten ihr Bild oder ihre Zählraten in die 4 m entfernte Wache (siehe Abb. 3.96):

- Laserscanner: reine Zählraten über serielles Kabel
- Zählkamera: HTML-Seite mit Zählraten über Power-over-Ethernet-Kabel (PoE)
- Stereokamera (Nr. 4): unkomprimierte Bilddaten zweier Perspektiven über IEEE-1394b-Kabel
- HDV-Kamera (Nr. 7): MPEG-2-komprimierten Videostrom über HDMI-Kabel



Abb. 3.94: (a) Zähltor mit den angebrachten Erfassungssystemen am Haupteingang des Forschungszentrum Jülich; (b) Bild der HDV-Kamera zur manuellen Überprüfung der automatischen Zählsysteme

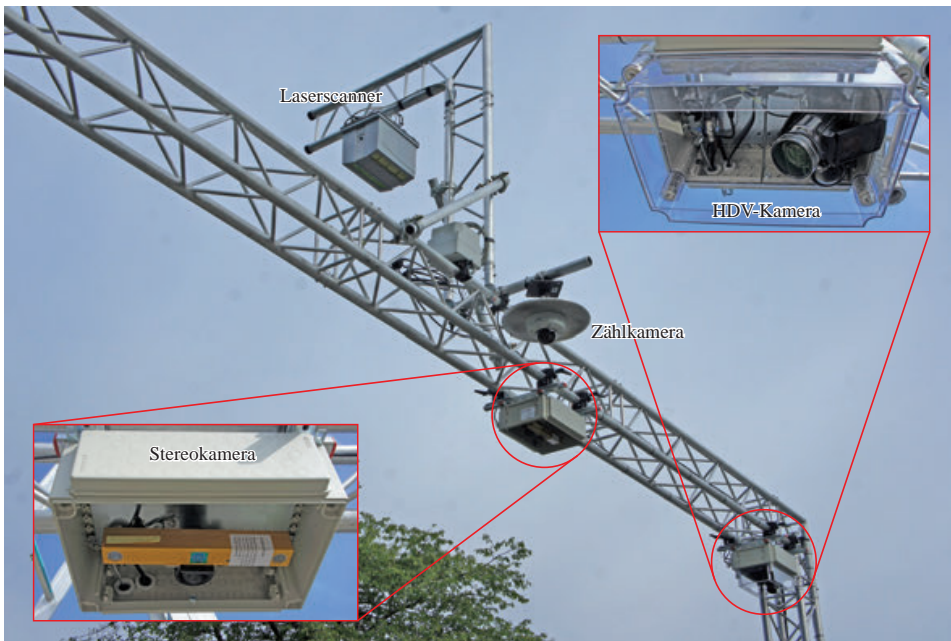


Abb. 3.95: Erfassungssysteme am Zähltor: Laserscanner des Produktes TrueView People Counter [513], Zählkamera des Produktes MoPeCo [515], Stereokamera für das selbst entwickelte Verfahren und HDV-Kamera zur Überwachung

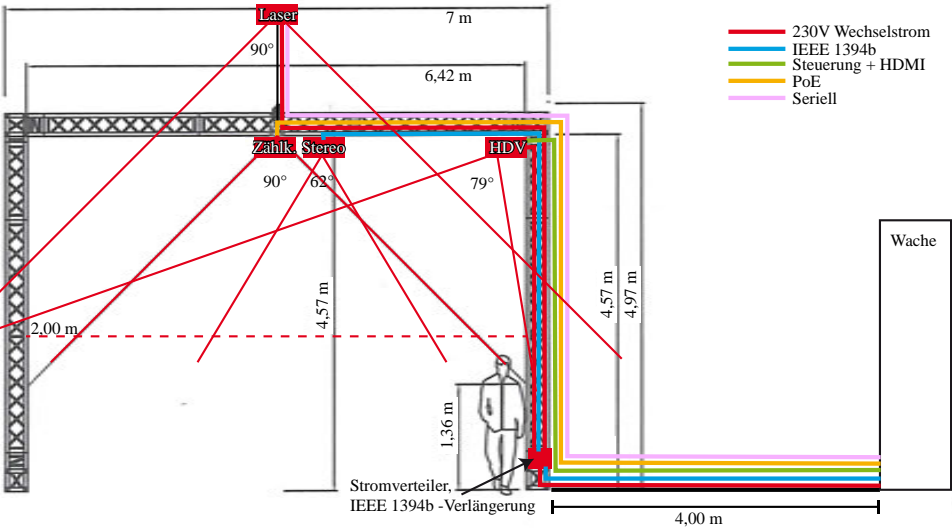


Abb. 3.96: Skizze der Vorderansicht des Zähltores am Haupteingang des Forschungszentrum Jülich zum „Tag der Neugier“; eingetragen ist neben dem Traversensystem die Lage der Erfassungssysteme mit ihren Erfassungsbereichen und die Kabelführung bis in die angrenzende Wache

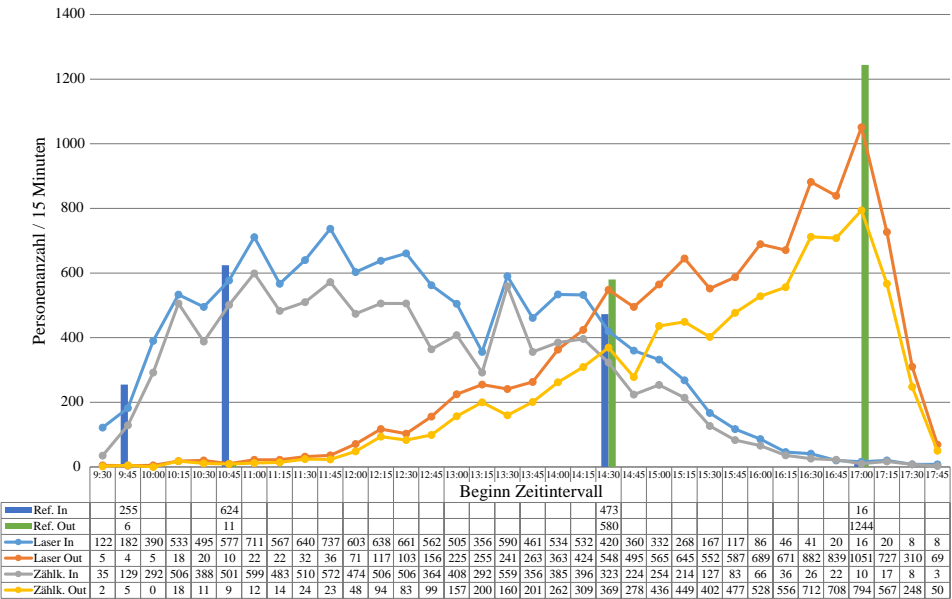


Abb. 3.97: Zähldaten der kommerziellen Erfassungssysteme; die ermittelte Anzahl liegt für den Ein- und Austritt beim Laserscanner zu jedem Zeitpunkt höher als bei der Zählkamera; die als Referenz dienenden manuell ermittelten Zählungen zu vier repräsentativen Zeitintervallen liegen noch über den Werten des Laserscanners

3.6.7.1 Vergleich der kommerziellen Systeme

Abb. 3.97 zeigt die Kurven der Zählraten beider kommerziellen Systeme getrennt nach ein- und ausgetretenen Personen. Das Stereosystem kann nicht im direkten Vergleich aufgetragen werden, da nicht der gesamte Durchgangsbereich eingesehen werden konnte. Für vier 15-Minuten-Intervalle wurden die Zählwerte von Hand erfasst und dienten als Referenzwerte. Die ausgewählten Intervalle bilden möglichst repräsentativ die unterschiedlichen Flüsse und Bewegungsrichtungen ab:

- 9:45 Uhr – 10:00 Uhr: geringer Fluss auf das Gelände des Forschungszentrum Jülich,
- 10:45 Uhr – 11:00 Uhr: mittlerer Fluss auf das Gelände,
- 14:30 Uhr – 14:45 Uhr: höchster bidirektionaler Fluss,
- 17:00 Uhr – 17:15 Uhr: höchster Fluss vom Gelände.

Positiv auf die Zählung werden sich folgende Faktoren ausgewirkt haben (die Punkte wiegen für beide Systeme unterschiedlich):

- das Wetter war die ganze Zeit über sonnig,
- Durchführung tagsüber und damit keine Dunkelheit,
- der globale Maximalfluss war gering: mit einer Durchgangsbreite von 6,42 m und maximal 1260 Personen in 15 Minuten ergibt sich ein Fluss von 1,4 Personen/s bzw. ein spezifischer Fluss von 0,22 Personen/(m s)),
- schweres Gepäck war selten.

Probleme werden folgende Umstände bereitet haben:

- Fahrradfahrer fahrend und schiebend,
- große Anzahl an Kindern, die oft an der Hand gehalten wurden oder sogar im Kinderwagen saßen,
- Mitführung von Luftballons durch 8 von 1250 Personen zwischen 17:00 Uhr und 17:15 Uhr,
- trotz geringen Gesamtflusses durch gruppenweises Eintreffen und Eintreten lokal höhere Dichten.

Die Tendenz beider Zählsysteme zur Unterschätzung der Personenzahl sind identisch. Die Abweichung zu den von Hand bestimmten Werten sind signifikant. Sie beträgt bei der monokularen Zählkamera im Durchschnitt ca. 35% (20% – 48%) und beim Laserscanner-System ca. 15% (8% – 29%) und schwankt stark für die ausgewerteten Intervalle.

Eine Verbesserung der Zählergebnisse dürften folgende Anpassungen bewirken:

- höhere Aufhängung (insbesondere der Zählkamera) und dadurch mögliche Einschränkung des Zählwinkels,
- keine Ablehnung von Objekten kleiner als 120 cm ggf. unter Zunahme der Falsch-Ermittlungs-Rate (für Laserscanner-System).

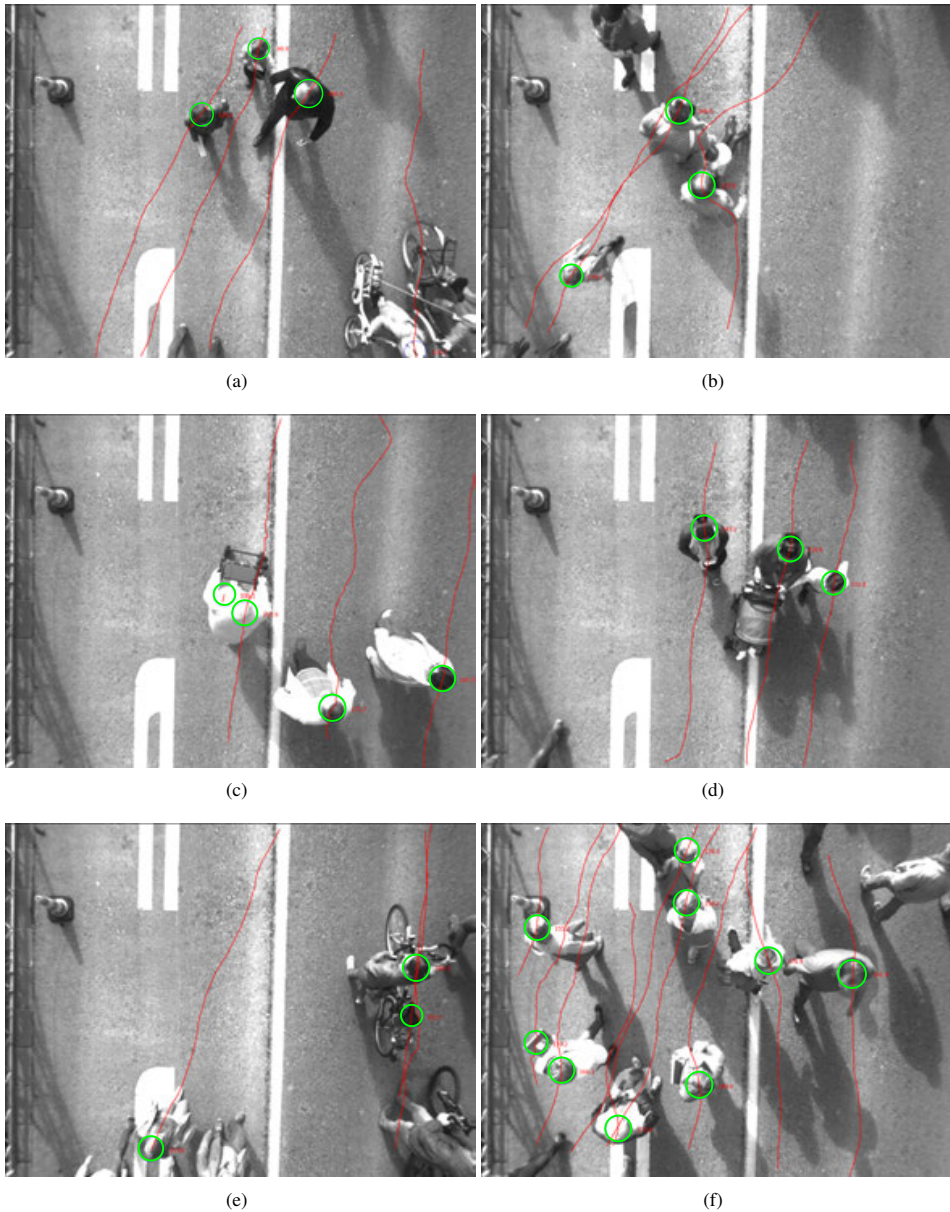


Abb. 3.98: Einzelbilder der Stereokamera mit eingezeichneten Laufwegen, detektierten Personengrößen und zumeist kritischen Szenen: (a, b) an der Hand genommene Kinder werden je nach temporärem Abstand zu den Eltern erkannt; (c) durch den dunklen Hintergrund und eine automatische Belichtungskontrolle kam es bei heller Kleidung oder Haaren zur Überbelichtung und zu Fehldetektionen; (d) der Kinderwagen wird zurecht nicht erkannt, aber auch nicht das darin befindliche Kind; (e) der Fahrradsattel wird als Person gezählt; (f) Bild der höchsten globalen Dichte während des Einsatztages und fälschlich als Person detektierte Zeitschriften

3.6.7.2 Selbstentwickeltes Verfahren

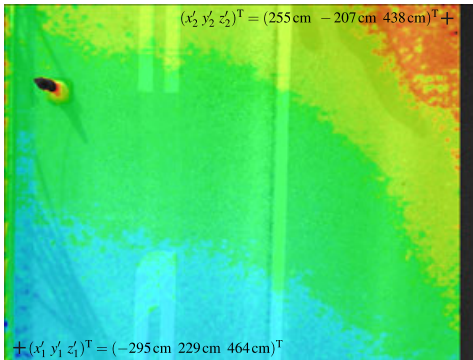


Abb. 3.99: Farbkodiert überlagerte Disparität des Hintergrunds zur Veranschaulichung der Schiefelage der Kamera; 3D-Koordinaten für Pixel in der linken unteren und rechten oberen Ecke zur Berechnung der Achsendrehung

Für die Stereokamera und Nutzung des selbstentwickelten Verfahrens wurde das Ergebnis für das Zeitintervall 14:30 Uhr – 14:45 Uhr untersucht. Es konnten aufgrund des in durchschnittlicher Kopfhöhe auf $338\text{ cm} \times 278\text{ cm}$ beschränkten Erfassungsbereichs nur 712 der 1053 Personen detektiert werden:

$$2(457\text{ cm} - 176\text{ cm}) \tan 31^\circ = 338\text{ cm}.$$

Der angewendete Blickwinkel der horizontalen Achse von 31° berücksichtigt bereits die fehlenden Disparitätswerte aufgrund sich nicht überlappender Sichtbereiche beider Perspektiven des Stereokamerasystems (siehe z. B. fehlende farbige Pixel im rechten Bildausschnitt von Abb. 3.39 und 3.99). Aus diesem Grund stimmt das Seitenverhältnis des Abdeckungsbereiches auch nicht

mit dem des Sensors der Kamera überein (siehe Tab. 4.2). Der Anteil der erkannten Personen von allen durch das Zähltor laufenden Personen von $712/1053 \approx 68\%$ ist größer als die abgedeckte Fläche von $338\text{ cm}/642\text{ cm} \approx 53\%$, da mehr Personen in der Mitte des Tores hindurchgelaufen sind.

Die Kamera konnte aufgrund der zeitlichen Rahmenbedingungen nicht genau senkrecht ausgerichtet werden. Die optische Achse lag

$$\arctan \frac{|z'_2 - z'_1|}{\|(x'_2 y'_2)^T - (x'_1 y'_1)^T\|} \approx 2^\circ$$

aus dem Lot (siehe Abb. 3.99).

Es kam in folgenden Situationen zu Fehlern durch zu viel gezählte Personen (falsch positive):

- 6 Fahrradsättel (siehe Abb. 3.98 e),
- 1 Fahrradkorb,
- 1 Luftballon,
- 2 erhobene Arme u. a. mit Zeitschriften (siehe Abb. 3.98 f),
- 4 Überbelichtungen (siehe Abb. 3.98 c).

Folgende Situationen führten zu einer Unterschätzung (falsch negative):

- 1 Erwachsener, da Kind auf den Schultern saß,
- 4 Kinderwagen mit Kindern (siehe Abb. 3.98 d),
- 1 Kind an der Hand eines Erwachsenen (siehe Abb. 3.98 b),

- 1 Überbelichtung.

Hingegen wurden andere an der Hand geführte Kinder (siehe Abb. 3.98 a) und Kinder, die sich im Kinderwagen nach vorne beugten, erkannt. Fahrradfahrer wurden auch detektiert, da der Wertebereich der Disparität so eingestellt wurde, dass Korrespondenzen bis zu einem Pixelabstand von 99 px gesucht wurden und damit gemäß Gleichung 3.23 bei einer durchschnittlichen Bodenentfernung von 455 cm Erhebungen bis

$$455 \text{ cm} - \frac{24 \text{ cm} \cdot 983 \text{ px}}{99 \text{ px}} = 218 \text{ cm}$$

bestimmt wurden.

Die Überbelichtung (u. a. Abb. 3.98 c) kam dadurch zustande, dass zum einen die Belichtungskontrolle automatisch durchgeführt wurde, damit das System unbeaufsichtigt den ganzen Tag über laufen konnte und bei Sonnenschein wie auch bewölktem Himmel kontrastreiche Bilder lieferte. Zum anderen war der Hintergrund der Aufnahme mit der asphaltierten Straße sehr dunkel, so dass bei der Automatik, die ein durchschnittliches Helligkeitsspektrum anstrebt, die Belichtungszeit für eine gute Abbildung des Vordergrundes zu lang eingestellt war und damit ab einem Helligkeitwert, z. B. hervorgerufen durch helle Kleidung oder Haare, alle Pixel auf Weiß abgebildet wurden. Dies führte zu fehlerhaften oder nicht bestimmbar Disparitäten und damit zu einem in dieser Region undefinierten Höhenfeld, woraus falsch positive und falsch negative Detektionen resultierten.

Abb. 3.98 f zeigt eine Szene mit der höchsten beobachteten globalen Dichte während des Einsatztages. Zu lokal höheren Dichten kam es nur, wenn Gruppen von zusammengehörenden Personen gemeinsam den Erfassungsbereich passierten.

Als Zählwert ergibt sich unter Berücksichtigung der Falsch-Ermittlungs- und Falsch-Negativ-Rate nach Zusammenlegung eine Überschätzung von 7 Personen (ca. 1%) und damit eine wesentlich bessere Abschätzung der Personenzahl. Zudem liefert das Verfahren nicht nur einen Zählwert, sondern bestimmt darüber hinaus Laufwege jeder einzelnen Person. Auch wird über das Höhenfeld die Personengröße bestimmt, die ggf. für eine Besucherstatistik herangezogen werden könnte.

Hingegen spricht für die kommerziellen Produkte, dass sie echtzeitfähig sind, einen größeren Abdeckungsbereich besitzen und zur Wahrung der allgemeinen Persönlichkeitsrechte (Recht am eigenen Bild) [516] reine Zählraten übertragen. Die Echtzeitfähigkeit stand bei der Entwicklung des eigenen Verfahrens nicht im Vordergrund (siehe Abschnitt 3.6.8), da die Bilddaten bei den Versuchen sowieso mit erfasst werden sollten, um diese zur qualitativen Analyse und manuellen Korrektur der Laufwege zur Verfügung stehen zu haben. Sollte eine Optimierung des Verfahrens erfolgen und es damit echtzeitfähig werden, so könnten die Bilddaten direkt nach der Auswertung gelöscht werden und damit die Persönlichkeitsrechte gewahrt bleiben. Wird zusätzlich die Recheneinheit an die Kamera gebracht, könnten auch nur noch anonyme Zählraten übertragen werden.

3.6.8 Rechenzeit

Zwar sind alle Verfahren nicht auf Performance hin optimiert worden bzw. stand bei der Auswahl von eingesetzten Methoden die Geschwindigkeit nicht im Vordergrund, trotzdem soll u. a. zur Verdeutlichung der relativen Geschwindigkeiten der einzelnen Schritte und Verfahren zueinander einige Rechenzeiten angegeben werden. Diese sind Tab. 3.10 zu entnehmen.

Die Performance war deshalb bei der Entwicklung nicht entscheidend, da für eine manuelle Sichtung bzw. qualitative Analyse von Versuchen oder einzelnen Phänomenen, die in den Laufwegen beob-

Versuch	V. reihe ¹	Kamera ²	Marker ³	Pers. zahl ⁴	Abb.	Vorvera.	Kalib.	Merkm.	Segment.	Objekts.	Verfolg.	gesamt (f/s)
Engstelle	3	1	1	27	3.2	4 (2%)	75 (41%)	0 (0%)	66 (36%)	3 (2%)	35 (19%)	183 (5,5)
DTU			1	(65)	3.100	0 (0%)	0 (0%)	0 (0%)	53868 (99%)	342 (1%)	170 (0%)	54380 (0)
SL-Rasch			1	211	3.101	0 (0%)	0 (0%)	11 (0%)	2004 (72%)	218 (8%)	570 (20%)	2802 (0,4)
Korridor	6	4	2	30	3.36	0 (0%)	23 (7%)	232 (69%)	31 (9%)	1 (0%)	51 (15%)	338 (3,0)
Tribüne	6	4	2	70	3.3	0 (0%)	23 (4%)	375 (59%)	86 (13%)	4 (1%)	150 (24%)	638 (1,6)
Raumeintritt	8		3	25	3.8 3	10 (3%)	210 (54%)	1 (0%)	100 (26%)	4 (1%)	75 (19%)	391 (2,6)
Korridor	8		4	21	3.9 4	0 (0%)	210 (91%)	16 (7%)	1 (0%)	0 (%)	4 (2%)	231 (4,3)
Feldstudie		4		4	3.98	0 (0%)	23 (3%)	586 (73%)	185 (23%)	3 (0%)	10 (1%)	807 (1,2)

Tab. 3.10: Durchschnittlicher anteiliger Zeitbedarf pro Frame in Millisekunden auf einem Kern eines Intel Core i7-2640M-Prozessors während des stationären Zustandes des Experiments bei maximaler Anzahl an Personen in dieser Versuchsreihe; ¹Versuchsreihe gemäß Tab. 4.1; ²Kameramodell gemäß Tab. 4.2; ³Marker gemäß Abb. 3.8 und 3.9; ⁴Anzahl der erkannten und verfolgten Personen zwischen zwei aufeinander folgenden Frames (nicht Gesamtzahl der Teilnehmer des Experiments); die Spalte „Abbildung“ verweist auf ein Bild des zugrundeliegenden Versuchs; folgende Spalten sind die nacheinander durchzuführenden Schritte zur Extraktion der Trajektorien und unterscheiden sich je nach Erkennungsmethode erheblich voneinander: die **Vorverarbeitung** enthält die Anwendung globaler Filter nebst Hinzugabe eines Rahmens, die **Kalibrierung** enthält ggf. die Entzerrung oder die Rektifizierung, folgende drei Spalten gehören zum Prozess der Erkennung: die **Merkmalsbestimmung** enthält die Hintergrundsubtraktion, die Bestimmung der Helligkeitsstufen, der Farbmaste oder der Disparitätsmatrix, die **Segmentierung** enthält die Bestimmung der Isolinien und deren Approximation durch Ellipsen, die **Objektsynthese** umfasst Methoden zur Personenidentifikation und die **Verfolgung** ist das reine Tracken aller detektierten Personen

achtet wurden, die originalen Videoaufnahmen erhalten bleiben sollten und damit ein nachgelagertes Extrahieren der Laufwege nicht in Echtzeit durchgeführt werden musste. Im gesamten Prozess der Datenerhebung, Extraktion und Analyse nimmt zudem die automatische Ermittlung der Laufwege einen nur geringen zeitlichen Raum ein.

Ist jedoch eine Echtzeitanalyse wichtig, da die Daten zur direkten Weiterverarbeitung, z. B. in einem Sicherheitssystem, benötigt werden oder aber der Zeitraum der Beobachtung insbesondere bei Feldstudien so lang ist, dass die Datenmenge der Videos zur Speicherung zu groß wird, müssen die Verfahren eine Bildwiederholrate erzielen, die mindestens der des Bilderfassungssystems (typischerweise zwischen 15 f/s und 30 f/s, vgl. Tab. 4.2) entspricht.

Neben einem Eingriff in die vorgestellten Verfahren zur Optimierung der Rechengeschwindigkeit wäre eine Steigerung der Geschwindigkeit leicht durch eine Parallelisierung erzielbar. Die Detektion arbeitet nur intraframe und nutzt somit keine Interframe-Informationen aus, so dass Einzelbilder für die Erkennung auf mehrere Prozessoren auch bei getrenntem Hauptspeicher verteilt werden könnten. Bei Systemen mit gemeinsamem Hauptspeicher könnten beim Tracking die zu verfolgenden Personen auf mehrere Prozessoren verteilt werden. Der Schritt der Kalibrierung könnte am Rechner wesentlich vereinfacht werden, wenn bei intelligenten Kameras wie in Fotoapparaten üblich an der Kamera selber die Entzerrung durchgeführt würde. Außer bei der Kalibrierung wäre zudem eine durchgehende Bildaufteilung (Überlagerung der Gebiete je nach Größe der Strukturelemente z. B. bei morphologischen Operationen) auf Prozessoren möglich, bei der im Randbereich wie bei überlappenden Kameras (siehe Abschnitt 4.1.1.2 und 3.3.5) die Laufwege weitgereicht bzw. verbunden würden. Viele der eingesetzten Techniken sind zudem sehr gut auf Grafikkarten portierbar, z. B. gibt es für die Bestimmung der Disparitätsmatrix, die bei den Verfahren mit Nutzung von Stereokameras den größten Zeitbedarf besitzt (siehe Spalte Merkmalsbestimmung in Tab. 3.10), freie Implementationen, u. a. vom BM-Ver-

fahren in der neusten Version (2.4.9) der Bibliothek OpenCV [446, 447] oder [517–520]. Parallele Implementationen oder Realisierungen für Grafikprozessoren (GPU, *graphics processing unit*) existieren auch bereits von in Abschnitt 2.3.1 vorgestellten allgemeinen Detektionsmethoden [521, 522].

Tab. 3.10 zeigt eine Auswahl von Versuchen, die unterschiedliche Verfahren nutzen oder Probleme bzw. Herausforderungen der Verfahren aufzeigen. Dabei sind auch zwei bisher noch nicht erwähnte Experimente von einer anderen Universität und einer Firma, die ebenfalls die Software *PeTrack* nutzen, aufgeführt, um Extrembeispiele der Laufzeit und der gleichzeitig in einem Frame vorkommenden Personen vorzustellen. Die generelle Verbreitung und Nutzung der Software *PeTrack* von anderen Gruppen kann nur anhand der registrierten Download-Zahl von 553 (Stand: 31.12.2014) erahnt werden, da die Software frei zum Download angeboten wird.

Die Versuchsreihe der Technischen Universität Dänemarks (DTU, *Danmarks Tekniske Universitet*) hat gezeigt, dass die Nutzung von gerichteten Isolinien gleicher Helligkeit bei stark texturierten Oberflächen durch die sehr große Zahl an Isolinien bei der Performance einbricht und zudem die sonst bei komplexen Markern geringe Falsch-Ermittlungs-Rate signifikant ansteigt. Um diesem Problem zu begegnen, könnte die für die lange Rechenzeit verantwortliche, starke Texturierung des Untergrundes in einer Reithalle (siehe Abb. 3.100) vermieden, der Bereich, in dem Personen anhand der getragenen Marker gesucht werden, eingegrenzt oder das jeweilige Binärbild der Helligkeitsstufen wie beim Farbmarker (siehe Abschnitt 3.2.2) von kleinen Störungen mittels morphologischen Operationen wie dem Öffnen und Schließen beseitigt werden.

Die auch in Ihrer Größe beeindruckende Versuchsreihe der Firma SL-Rasch zur Untersuchung des Einflusses von Säulen in einem 5 m breiten Korridor zeichnete sich dadurch aus, dass sehr viele Personen pro Bild erkannt werden mussten (siehe Abb. 3.101).

Die einzelnen in der Tab. 3.10 aufgeführten Schritte unterscheiden sich je nach Verfahren insbesondere bei der Erkennung erheblich voneinander. Zum Erkennungsschritt gehören die Spalten Merkmalsbestimmung, Segmentierung und Objektsynthese. Die Unterteilung im Erkennungsschritt orientiert sich an der Prozesskette zur Informationsextraktion in der Bildverarbeitung [30]. Die einzelnen Spalten umfassen:

Vorverarbeitung: die Anwendung globaler Filter (z. B. Helligkeit, Kontrast, Bildspiegelung) nebst Hinzugabe eines Rahmens (Abschnitt 3.1.2); das Einlesen der Bild- oder Videodaten wurde nicht berücksichtigt, da die Daten in unterschiedlichen Codecs vorliegen und an verschiedenen Speicherorten (Flash-Speicher, interne und externe Festplatten und Datenserver) abgelegt sind,

Kalibrierung: als rechenintensive Funktionen die Entzerrung (Abschnitt 3.1.2) oder Rektifizierung (Abschnitt 3.4.2),

Merkmalsbestimmung: als Vorverarbeitung der Merkmalsbestimmung die Hintergrundsubtraktion (Abschnitt 3.2 und 3.5.1) und die Bestimmung der Helligkeitsstufen (Abschnitt 3.2.1), der Farbmärke (Abschnitt 3.2.2) oder der Disparitätsmatrix (Abschnitt 3.4.2) bzw. des Höhenfeldes,

Segmentierung: die Bestimmung der Isolinien (Abschnitt 3.2.1, 3.5.2) und deren Approximation durch Ellipsen (Abschnitt 3.5.3),

Objektsynthese: die Personenidentifikation, z. B. durch Lage und Größe der Ellipsen (Abschnitt 3.2.1 und 3.5.5),

Verfolgung: das reine Tracken aller detektierten Personen (Abschnitt 3.3).



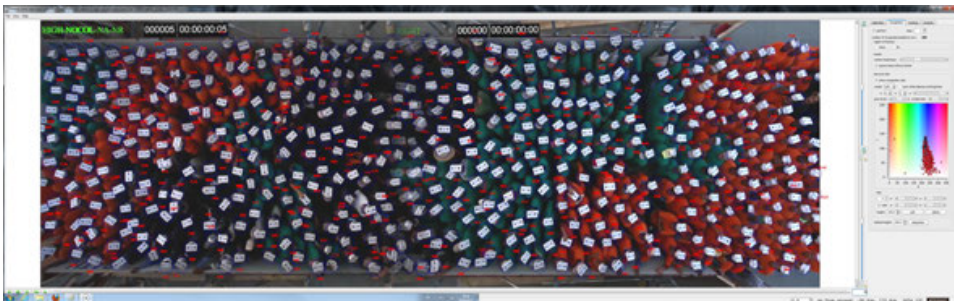
Abb. 3.100: Versuch der Technischen Universität Dänemarks (DTU) mit einer aufgrund des stark texturierten Untergrundes sehr langen Laufzeit der Extraktion der Laufwege; Bild von Lizette Petersen, DTU [523,524]



(a)



(b)



(c)

Abb. 3.101: Versuchsaufbau, Schrägsicht und Aufsicht einer Versuchsreihe der Firma SL-Rasch zur Untersuchung des Einflusses von Säulen in einem 5 m breiten Korridor; (c) zusammengesetztes Bild der beiden senkrecht ausgerichteten Kameras zur Extraktion der Laufwege mittels PeTrack; Bilder erstellt von SL-Rasch

Die Zeiten sind in Millisekunden angegeben und wurden auf einem Laptop (Lenovo ThinkPad T420s) mit folgenden Kennzahlen bestimmt:

Prozessor: Intel Core i7-2640M, Basistakt 2,8 Ghz,

Hauptspeicher: 8 GB,

Betriebssystem: Windows 7, 64 bit.

Dabei wurde nur ein Thread auf einem Rechenkern von zwei Kernen mit vier parallel möglichen Threads (Hyperthreading) genutzt.

Bei der Vorverarbeitung wird nur dann merklich Rechenzeit benötigt, wenn aufgrund von negativen radialen Parametern zur Entzerrung im Verarbeitungsschritt der Kalibrierung der relevante Bildinhalt über den Bildrand hinaus geht und dafür jedem Frame ein Rahmen hinzugefügt werden soll (siehe Abschnitt 3.1.2).

Die Kalibrierung konnte nur bei Bilddaten durchgeführt werden, die selbst erzeugt wurden und somit entweder die benötigten Parameter über Kalibrationsmuster bestimmt werden konnten (siehe Abschnitt 3.1.2) oder aber wie bei den Stereokameras vom Kamerahersteller vorgegeben waren. Das Verfahren zur Rektifizierung der beiden Bilder der Stereokamera arbeitet schneller als die Entzerrung der monokularen Kameras.

Die Laufzeit der Merkmalsbestimmung ist für die Marker mit Strukturelementen, für die keine Disparitätsmatrix zur Größenbestimmung berechnet werden muss oder kann (Marker Nr. 1 und 3), zu vernachlässigen. Die Rechenzeit zur Bestimmung der Schwellwertbilder der Helligkeitsstufen steigt linear in Abhängigkeit von der Pixelzahl des interessierenden Bereichs (ROI) und der Anzahl der Stufen. Bei den Experimenten der Firma SL-Rasch liegt diese Zeit etwas höher, da die Anzahl an Schwellwerten höher angesetzt werden musste, weil die Helligkeit im Bild sehr unterschiedlich ausfiel. Beim Farbmarker muss der Farbraum gewechselt und morphologische Operationen durchgeführt werden, was die Merkmalsbestimmung etwas aufwändiger ausfallen lässt. Die mit Abstand langsamste Merkmalsbestimmung besitzen Verfahren, die als Merkmal die Disparitätswerte bzw. die Höheninformationen benötigen, sei es zur Größenbestimmung oder aber zur Bildung von Höhenlinien zur markerlosen Personendetektion. Die Laufzeit der Berechnung der Disparitäts-Matrix ist dabei auch bei gleich großen Ausgangsbildern nicht identisch, sondern ist z. B. für Tribünen-Experimente länger als für Korridor-Experimente, da der Wertebereich der möglichen Disparitätswerte größer ist und somit Punktkorrespondenzen in einem größeren Intervall gesucht werden müssen (siehe Abschnitt 3.4.2).

Bei der Segmentierung ist die Erzeugung der Beschreibung der Segmente bei den Markern mit Strukturelementen mit gerichteten Isolinien gleicher Helligkeit am aufwändigsten. Bei einer sehr großen Zahl von Segmenten kann dieser Schritt wie bei den DTU-Experimenten das Verfahren fast zum Erliegen bringen.

Die Objektsynthese besitzt geringe Laufzeiten, auch bei der aufwändigsten markerlosen Erkennung, die die pyramidalen Ellipsenstapel nach der Form eines menschlichen Oberkörpers durchsuchen. Die Suche nach Objekten dauert für die Fälle im Verhältnis zur Anzahl an erkannten Personen lang, in denen viele Segmente zur Synthese weitergeleitet werden, die keine Person darstellen, wie es insbesondere bei den DTU-Experimenten, aber auch den Experimenten von SL-Rasch der Fall war.

Dass die Laufzeit des Trackings nicht linear mit der Anzahl der zu verfolgenden Personen ansteigt, liegt zum einen an der unterschiedlichen Anzahl an Pixeln, die durch einen Kopf abgedeckt werden,

und damit unterschiedlich großen Bereichen dieser Größenordnung, die in Folgebildern auf Gleichheit überprüft werden müssen, und zum anderen an der Anzahl an Ebenen in den Gauß-Pyramiden. Beim Farbmarker wurde nur eine Ebene und bei den Markern mit kleinen zentralen Strukturelementen, wie in Abschnitt 3.3 beschrieben, vier Ebenen für ein exaktes und stabiles Tracking verwandt. Beim Korridor-Experiment der Versuchsreihe Nr. 8 sind es z. B. nur ca. 400 Pixel und eine Ebene, beim Engstellenversuch der Versuchsreihe Nr. 3 ca. 2000 Pixel und vier Ebenen.

Bei den Experimenten wurde innerhalb der Versuchsreihe das Experiment mit der maximalen Personenzahl gewählt und die Zeitmessung während des stationären Zustandes durchgeführt. Zudem wurde die ROI auf den Bereich der Versuchsdurchführung eingeschränkt. Dabei hängt die Dauer der einzelnen Schritte nicht nur vom Verfahren und von der Größe der ROI ab, sondern jeder Schritt ist abhängig von vielen weiteren Faktoren, die sich überdies unterschiedlich stark in Abhängigkeit vom verwendeten Verfahren auf die einzelnen Schritte auswirken. Diese Faktoren sind neben der Pixelzahl der ROI z. B. die Gesamtpixelzahl, die Personenzahl, die Verwendung von Filtern (Helligkeit, Farbraumen, Hintergrundsubtraktion), die Farbverteilung, der Kontrast und die Textur im Bild und die Parameter der einzelnen Schritte wie die Anzahl der Helligkeitsstufen oder die Größe der Vergleichsmaske bei der Bestimmung der Disparität.

Aussagen zu Laufzeiten der unterschiedlichen in PeTrack integrierten Verfahren zur Bestimmung der Disparitäts-Matrix können Abschnitt 3.4.2 entnommen werden.

3.6.9 Fehlereinfluss auf Maße der Fußgängerdynamik

Fehler in den Laufwegen können unterschiedliche Arten von Ungenauigkeiten aufweisen:

1. Fehler bei der Erkennung: quantifizierbar über die Falsch-Negativ- und Falsch-Ermittlungs-Rate,
2. systematischer Fehler der Laufwege in Abhängigkeit vom Abstand zur Bildmitte: z. B. Anstieg
 - (a) des Fehlers der ermittelten Position (u. a. beim Farbmarker und der markerlosen Erkennung),
 - (b) der Fehlervarianz der ermittelten Position (u. a. aufgrund nicht beachteter, fehlerhafter oder abgeschätzter Personengrößen),
3. lokale Fehler innerhalb der Laufwege: z. B. Genauigkeit in der Bestimmung der gleichen Position auf dem Kopf (u. a. bei der markerlosen Erkennung)

Zur Untersuchung des Einflusses dieser Fehler wurden ausgehend von höhenkorrigierten Laufwegen, die mittels eines strukturierten Markers bestimmt wurden, diese Fehler exemplarisch zum Teil künstlich herbeigeführt. Folgende Abschnitte zeigen den Einfluss dieser Fehler auf Messgrößen in der Fußgängerdynamik.

3.6.9.1 Fehler bei der Erkennung

Der Einfluss einer fehlerhaften Detektion von Personen ist in Abb. 3.102 und 3.103 deutlich erkennbar. Ausgehend von einem optimalen Datensatz eines Versuches an einer T-Kreuzung aus Versuchsreihe Nr. 3 wurden 2,5%, 5%, 10%, 20% und 40% der Laufwege zufällig eliminiert, was der gleichen

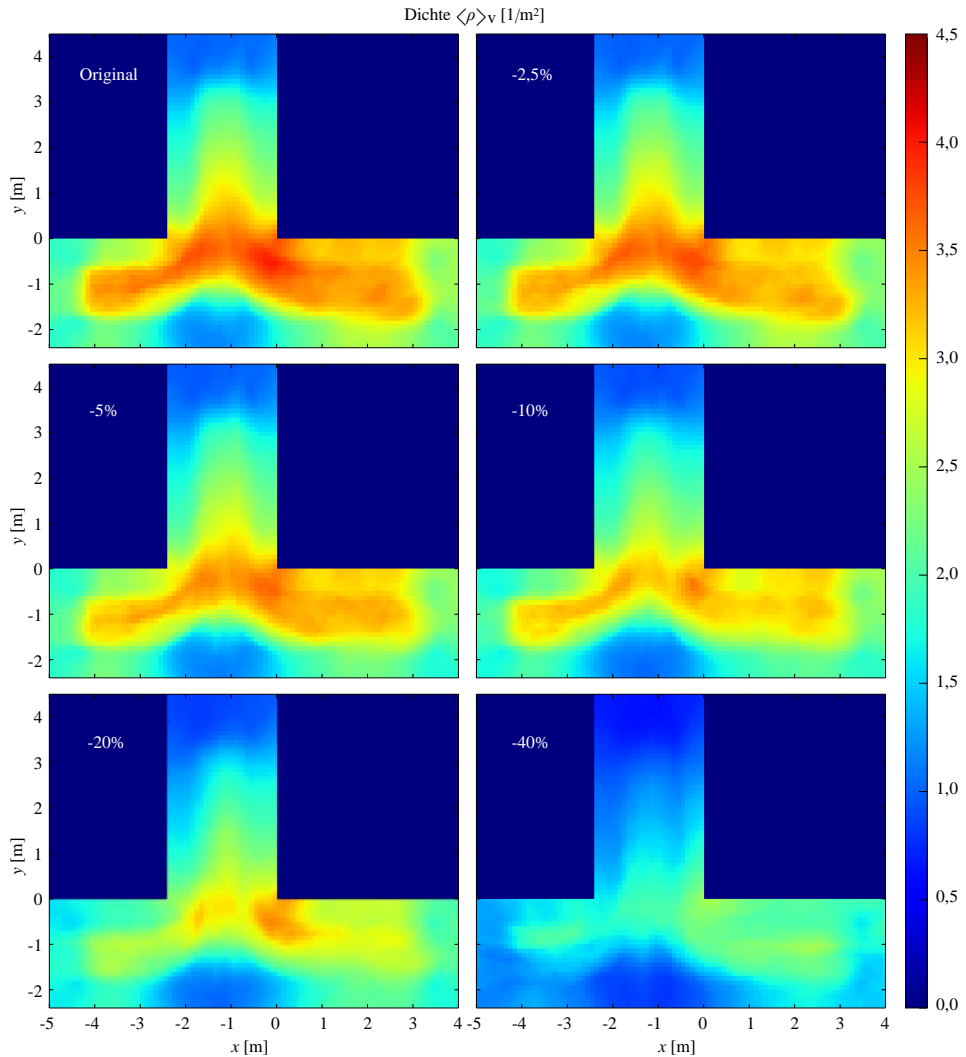


Abb. 3.102: Dichte-Profil mit Kacheln der Größe $10\text{ cm} \times 10\text{ cm}$ eines Experimentes an einer T-Kreuzung der Versuchsreihe Nr. 6 mit 303 Versuchsteilnehmern, von denen zufällig 2,5%, 5%, 10%, 20% und 40% zur Simulation von Falsch-Negativ-Raten gelöscht wurden (vgl. Abschnitt 4.4.2); Verteilung der Dichte in Abb. 3.103

Falsch-Negativ-Rate entspricht. Die Dichte-Profile (siehe Abschnitt 4.3) in Abb. 3.102 zeigen die Veränderung der Dichte-Verteilung innerhalb der Versuchsgeometrie. Die Dichte nimmt ausgehend vom korrekten Durchschnittswert von $2,47/\text{m}^2 \pm 0,76/\text{m}^2$ kontinuierlich mit $2,40/\text{m}^2 \pm 0,73/\text{m}^2$, $2,31/\text{m}^2 \pm 0,70/\text{m}^2$, $2,19/\text{m}^2 \pm 0,68/\text{m}^2$, $2,00/\text{m}^2 \pm 0,61/\text{m}^2$ und $1,48/\text{m}^2 \pm 0,46/\text{m}^2$ ab.

Dieser sichtbare Einfluss auf die für die Sicherheit in einer Menschenmenge entscheidende Messgröße macht deutlich, wie wichtig verlässliche Trajektorien sind und warum neben Verfahren mit sehr

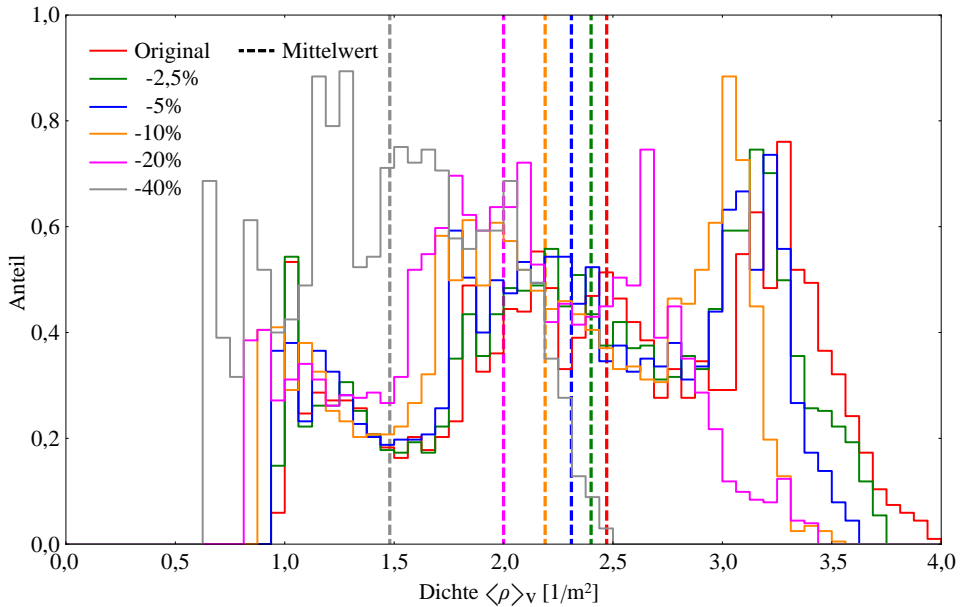


Abb. 3.103: Verteilung der Dichte bei fehlenden Trajektorien eines Versuchsdatensatzes (-2,5%, 5%, 10%, 20% und 40%); normiert, so dass jedes Integral über alle Dichten 1 ergibt

geringen Fehlerraten zudem eine manuelle Durchsicht durchgeführt wurde.

Für das oft zur Beschreibung der Fußgängerdynamik herangezogene Fundamentaldiagramm (siehe Abb. 1.4) zur Darstellung der Abhängigkeit der Geschwindigkeit von der Dichte führt die Verringerung der ermittelten Dichte dazu, dass der Abfall im Diagramm schneller als in Wirklichkeit verläuft, da die über die Laufwege ermittelte Geschwindigkeit unabhängig von der Fehlerrate der Detektion ist.

3.6.9.2 Ortsabhängiger Fehler

Der systematische, ortsabhängige Fehler der Laufwege ist bereits bei der Untersuchung der einzelnen Marker beschrieben worden (vgl. Abschnitt 3.6.4.3 und 3.6.6.2). Dabei wurde ebenfalls die Wirkung auf die Geschwindigkeit als eine wichtige Messgröße der Fußgängerdynamik dargelegt (siehe Abb. 3.79 und 3.91).

Seien α_x und α_y die vorzeichenbehafteten Blickwinkel zur optischen Achse entlang der jeweiligen Koordinatenachse, so ist der ungefähre Fehler für Versuchsaufbauten ähnlich der der ebenen Hermes-Experimente (Nr. 6) in Abhängigkeit zum Abstand zur Bildmitte \bar{c}' gemäß Gleichung 3.33 für den Farbmarker

$$e_c \approx 0,12 \text{ cm}/^\circ \begin{pmatrix} \alpha_x \\ \alpha_y \\ 1 \end{pmatrix}$$

und gemäß Gleichung 3.38 für das markerlose Verfahren

$$e_o \approx (p_i - \bar{c}')(0,02 - 0,006/^\circ \begin{pmatrix} \alpha_x \\ \alpha_y \\ 1 \end{pmatrix}).$$

Die Geschwindigkeit wird dadurch jeweils um durchschnittlich 1,2% bzw. 1,7% unterschätzt.

3.6.9.3 Ortsabhängige Fehlervarianz

Die Fehlerart 2 b einer ortsabhängigen Fehlervarianz (siehe Liste auf Seite 174) tritt z. B. aufgrund der Berücksichtigung bzw. Nichtberücksichtigung der Personengröße bei der Positionsbestimmung in der Bewegungsebene auf. Der mögliche Fehler in der ermittelten Position durch eine fehlerbehaftete oder vernachlässigte Personengröße steigt systematisch von der Bildmitte zum Bildrand bzw. zur Bildecke (siehe Abschnitt 3.4).

Die Auswirkung soll anhand einer Untersuchung von Nachbarschaftsbeziehungen bei einem Engstellenversuch mit einer Engstellenbreite von 200 cm aus Versuchsreihe Nr. 3 (siehe Abb. 3.2) gezeigt werden.

Für die Betrachtung von Nachbarschaftsbeziehungen in einer Personengruppe z. B. zur Analyse von Laufmustern von Gruppen ist der Abstand zwischen den Personen eine wichtige Größe. Abb. 3.104 zeigt zum einen alle Laufwege des Versuchs mit Berücksichtigung der Personengröße über farbko-dierte Größenklassen (rot) und Laufwege, für deren Bestimmung lediglich eine einheitliche Durchschnittspersonengröße verwandt wurde (grün). Die farbig entsprechenden Verteilungen des minimalen Personenabstandes zeigen für den Bereich direkt unterhalb der Kamera (gestrichelt eingefasster Bereich) eine nur geringfügige Vergrößerung des Abstandes ($39,5 \text{ cm} \pm 6,3 \text{ cm}$ im Gegensatz zu $39,4 \text{ cm} \pm 6,5 \text{ cm}$) bei Berücksichtigung der Personengröße, was auf eine etwas zu klein angenommene Durchschnittspersonengröße zurückzuführen ist. Der Personenabstand am Rand unterscheidet sich stärker und liegt mit Berücksichtigung der Personengröße im Durchschnitt bei $43,4 \text{ cm} \pm 7,9 \text{ cm}$ und ohne Berücksichtigung bei $43,2 \text{ cm} \pm 8,6 \text{ cm}$ und besitzt damit ohne Nutzung von Größenklassen insbesondere eine erhöhte Standardabweichung, was auf die ungenauere Positionsbestimmung zum Rand zurückzuführen ist. Diese ersichtliche Änderung der individuellen Abstände kann Einfluss auf die Bewertung der Nachbarschaftsbeziehungen haben.

Unabhängig von der Fehlerbetrachtung ist weiter zu beobachten, dass, wie auch in Abb. 4.19 zu sehen, der Personenabstand kurz vor der Engstelle kleiner und damit die Dichte größer als innerhalb der Engstelle ist.

3.6.9.4 Ungenauigkeit des Detektionsortes

Neben einem systematischen Fehler kann die ermittelte Position einer Person generell mit einer Ungenauigkeit behaftet sein bzw. ihr Laufweg lokale Fehler aufweisen, wenn das Verfahren nicht in jedem Frame den gleichen Ort auf dem Kopf als Position einer Person annimmt. Dies gilt insbesondere für das markerlose Verfahren und resultiert zum großen Teil aus der fehlerbehafteten Bestimmung der Disparitätsmatrix bzw. des Höhenfeldes (siehe Abschnitt 3.4.3).

Aufgrund dieser ungenauen Ortung sind die Laufwege der markerlosen Erkennung unruhig (siehe Abb. 3.92), so dass Details in der Struktur der Laufwege verloren gehen können.

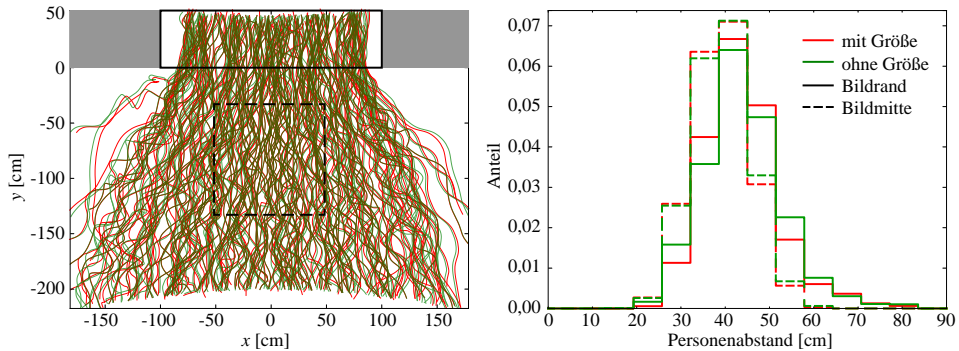


Abb. 3.104: Personenabstand in Abhängigkeit von der Berücksichtigung der Personengröße für einen Engstellenversuch mit einer Engstellenbreite von 200 cm aus Versuchsreihe Nr. 3 (siehe Abb. 3.2); die roten Pfade zeigen Trajektorien, bei denen die Personengröße berücksichtigt wurde, bei den grünen Pfaden wurde für die Bestimmung der Position in der Bewegungsebene eine Durchschnittspersonengröße angenommen; die Verteilung des Personenabstandes zeigt, dass für randständige Laufwege des durchgezogenen umrandeten Bereichs die Vernachlässigung der Personengröße die Standardabweichung des Abstandes vergrößert; direkt unter der Kamera vergrößert sich der gesamte Abstand nur leicht, was auf die Verwendung einer etwas zu großen Durchschnittspersonengröße zurückzuführen ist; der Personenabstand ist vor der Engstelle kleiner als in der Engstelle und damit die Dichte größer (vgl. Abb. 4.19); normiert, so dass jedes Integral über alle Personenabstände 1 ergibt

Der Einfluss der Ungenauigkeit soll anhand einer Untersuchung zur Linienbildung in einem Korridor gezeigt werden. Das Studium der Linienbildung dient u. a. dem Verständnis des maximalen Flusses bzw. der Kapazität einer Engstelle oder eines Ganges. Abb. 3.105 zeigt die Laufwege bei einem unidirektionalen Passieren eines 1 m breiten Korridors. Die Verteilung der Laufwege im Querschnitt wird durch Klassifizierung nach ihrer x-Koordinate der die schwarze Linie bei 2 m querenden Personen bestimmt. Die originalen Trajektorien (rot) zeigen in der entsprechend farbigen Verteilung der x-Werte eine klarere Ausprägung der existierenden Linien. Die korrekten Laufwege wurden durch eine normalverteilte Störung von $6\text{ cm} \pm 3\text{ cm}$ mit einem Fehler versehen (blau). Bei deren Verteilung der x-Werte im Querschnitt verringert sich die Ausprägung der Linien sichtlich. Bei für breitere Korridore weiter verschwimmenden Linien (vgl. [19]) kann die Ungenauigkeit in der Positionsbestimmung dazu führen, dass existente Linien sich nicht mehr im Querschnitt widerspiegeln.

3.6.10 Vergleich der Detektionsmethoden

In den letzten Abschnitten wurden einige Ergebnisse, insbesondere Fehler oder Probleme, der verschiedenen Verfahren zur Extraktion der Laufwege besprochen. Tab. 3.11 führt noch einmal die Ergebnisse zusammen und zeigt, dass trotz der vielen Einschränkungen, die in diesem Abschnitt 3.6 benannt wurden, die Fehlerraten für alle Verfahren sehr gering sind. Natürlich können diese Zahlen nur einen Anhaltspunkt geben, denn neben der reinen Erkennung und der Glattheit der Laufwege könnten viele andere Eigenschaften betrachtet werden, z. B. die Empfindlichkeit gegen Lichtänderungen oder Schrägstellung der Kamera, die Quote von Erkennung, reinem Tracking und Extrapolation, oder die Stabilität in der Erkennung zum Bildrand. Alle Ergebnisse sind zudem abhängig von vorgenommenen Einstellungen im Erkennungsprozess, u. a. der Wahl der interessierenden Region, in der die Detektion durchgeführt wird, die Regelmäßigkeit, in der die Erkennung neben dem reinen Tracking durchgeführt wird, die eingesetzten Methoden zur Bereinigung der Laufwege, wie der Elimination von nur

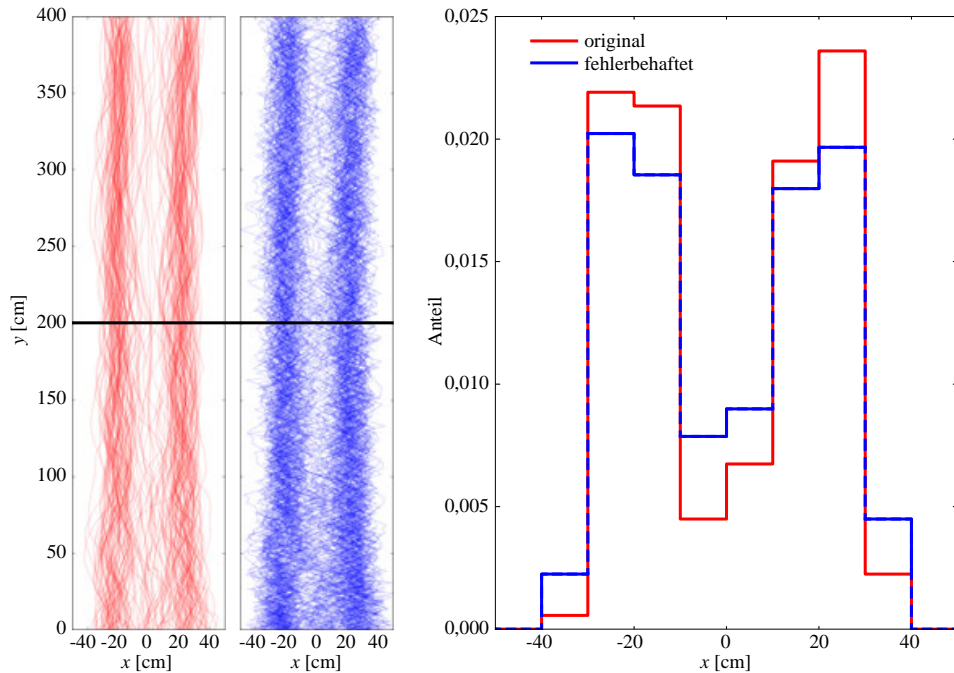


Abb. 3.105: Eine Ungenauigkeit in der Positionsbestimmung entlang der gesamten Trajektorien führt beim Studium der Linienbildung dazu, dass deren Ausprägung sich verringert; die originalen Laufwege (rot) zeigen eine klarere Aufteilung auf die Linien im Querschnitt als die mit einem normalverteilten Fehler versehenen Laufwege (blau)

kurzfristig erkannten Personen, und die einstellbaren Parameter der einzelnen Verfahren. Eine erschöpfende Analyse ist daher nahezu unmöglich, aber neben Tab. 3.11 gibt Tab. 3.1 einen Überblick über qualitative Eigenschaften der eingeführten Marker. Zusammenfassend einige Empfehlungen:

- der Farbmarder ist für Versuche mit mäßigen Dichten, keinen zu großen Unterschieden in der Personengröße und bei Interesse insbesondere an der Bewegung in Hauptbewegungsrichtung empfehlenswert; für den Farbmarder ist eine hohe Aufhängung der aufzeichnenden Kamera anzustreben, da er wie die markerlose Erkennung einen ortsabhängigen Fehler aufgrund der perspektivischen Verzerrung verursacht,
- präzise Laufwege sind über Marker mit Strukturelementen extrahierbar; die Wahl der Elemente muss sorgfältig ausgesucht werden und u. a. Dinge wie die Personendichte, Versuchsumgebung, Pixelabdeckung und Informationsgehalt berücksichtigen,
- eine markerlose Erkennung ist aus nahezu senkrechtem Blickwinkel auch bei hohen Dichten mit dem vorgestellten Verfahren möglich; für Schrägsichten wie bei Überwachungskameras üblich, müssen andere Verfahren herangezogen werden [33, 99, 105].

Eng damit verbunden sind vorteilhafte Bedingungen bei den technischen Komponenten sowie dem Aufbau und der Durchführung der Versuche für präzise Laufwege und deren mit wenig Fehlern behaftete Extraktion:

Versuch	V. reihe ¹	Kamera ²	Marker ³	Abb.	erkannte Personen	falsch positiv	falsch negativ	Beschleunigung [m/s ²]
Engstelle	3	1	1	3.2	177	0% (0)	0,6% (1)	0,7 ± 0,7
Korridor	3	1	1	4.2 a	45	0% (0)	0% (0)	0,5 ± 0,4
Korridor	6	4	2	3.36	79	3,8% (3)	0% (0)	0,7 ± 0,6
Tribüne	6	4	2	3.3	263	1,1% (3)	0% (0)	2,2 ± 2,7
T-Kreuzung	6	4	2	3.86 a	304	0,3% (1)	0% (0)	1,2 ± 0,7
T-Kreuzung	6	4	-	3.86 b	300	0% (0)	1,0% (3)	6,5 ± 6,8
T-Kreuzung	6	4	2+	3.92	303	0% (0)	0% (0)	1,2 ± 0,7
Ecke	6	4	-	3.44	162	0% (0)	2,4% (4)	3,6 ± 4,8
Raum	8	-	3	3.8 3	25	0% (0)	0% (0)	0,6 ± 0,5
Korridor	8	-	4	3.9 4	21	0% (0)	0% (0)	2,8 ± 1,8
Schlange	8	-	4	3.76	26	0% (0)	0% (0)	2,2 ± 1,7
Feldstudie	-	4	-	3.98	712	2,0% (14)	1,0% (7)	3,4 ± 11,0

Tab. 3.11: Erkannte Personen, Falsch-Ermittlungs- und Falsch-Negativ-Rate für alle vorgestellten Erkennungsmethoden; ¹Versuchsreihe gemäß Tab. 4.1, ²Kameramodell gemäß Tab. 4.2, ³Marker gemäß Abb. 3.8 und 3.9 („-“ verwendet das markerlose Verfahren, „2+“ kombiniert das markerlose Verfahren mit Marker Nr. 2); die Spalte „Abbildung“ verweist auf ein Bild des zugrundeliegenden Versuchs; der Betrag der Beschleunigung wurde gemäß Gleichung 3.44 bestimmt

- hohe Auflösung des Kamera-Sensors,
- große Brennweite des Objektivs,
- hohe Aufhängung der Kamera,
- hohe Bildwiederholrate,
- verlustlose Speicherung,
- viel diffuses Licht mit einheitlicher Farbtemperatur und matten, diffus spiegelnden Oberflächen,
- Umgebung mit viel Textur für das markerlose Verfahren und mit wenig Textur für auf Marker mit Strukturelementen basierende Verfahren.

Die Verbesserung einiger Bedingungen verschlechtert andere, so dass oft ein Kompromiss gefunden werden muss. Zum Beispiel verringert eine hohe Aufhängung die Pixelabdeckung oder eine große Brennweite den einsehbaren Versuchsbereich. Genauer werden diese Aspekte im folgenden Kapitel besprochen.

4

Versuche

Wie in der einführenden Motivation in Abschnitt 1.1 beschrieben, war es nötig, eigene Versuche durchzuführen, um verlässliche Daten für das Verständnis der Fußgängerdynamik zu erhalten und daraufhin Modelle entwickeln zu können. Zudem können die Modelle mit den erhobenen Daten kalibriert und validiert werden.

Seit 2005 wurden daher von uns eine Vielzahl von Versuchen durchgeführt (siehe Tab. 4.1), wobei die Komplexität der Szenarien anfangs (bis Versuchsreihe Nr. 5) geringer ((eindimensionale) Korridore, Engstellen) und die Flusstypen einfacher (unidirektional) waren und später auf komplexere Szenarien (Ecken, Treppen etc.) und Flusstypen (bidirektional, T-Kreuzungen, vereinigende Flüsse) ausgeweitet wurden. Die Vielzahl der Szenarien kann auch Abb. 4.1, 4.2 und 4.3 entnommen werden. Die Abbildungen zeigen einen Überblick an Kameraeinstellungen von verschiedenen Versuchen, aus denen die Laufwege gewonnen wurden. Trotzdem fehlen weiterhin wichtige Standardszenarien, von denen noch keine belastbaren Daten über die Bewegung der Fußgänger in diesen Geometrien vorliegen (z. B. Kreuzungen).

Die Hermes-Versuche (Versuchsreihe Nr. 6) waren hinsichtlich des Umfangs, der Komplexität und der Qualität der ermittelten Laufwege bereits weltweit einmalig, bei den BaSiGo-Versuchen (Nr. 9) wurde der Umfang und die Qualität noch weiter erhöht. In der neuen Versuchsreihe wird die Anzahl der Probanden merklich gesteigert, vor allem um besonders hohe bis hin zu kritischen Dichten herbeizuführen (siehe Video Nr. 5 und 9), und die Qualität der Laufwege dahingehend gesteigert, dass jede Person über ihren Marker eindeutig identifizierbar ist und somit die Daten aufgrund von detaillierten Fragebögen für eine Trajektorie umfangreicher werden (z. B. Alter, Fitness, Geschlecht, Kulturkreis, Zusammengehörigkeit).

Insbesondere in den letzten zehn Jahren wurden auch Versuche von anderen Forschergruppen durchgeführt, die die unterschiedlichsten Fragestellungen der Fußgängerdynamik zu beantworten versuchten. In den einzelnen Ländern sind dies ohne Anspruch auf Vollständigkeit für China [118–120, 140, 146, 147, 529, 530], in Frankreich im Rahmen des Pedigree Projektes [143, 156, 157, 325], in Japan [121–123, 531] und in Deutschland [115, 116].

Generell ist zur Untersuchung der Fußgängerdynamik eine genaue Studie derselben wichtig. Dabei ist die Abhängigkeit der Bewegung von verschiedensten Parametern zu quantifizieren. Um Einzelabhängigkeiten herausarbeiten zu können, müssen Versuche unter Laborbedingungen (*ex situ*) durchgeführt

Nr.	Jahr	Ort	Pers. anz.	Vers. anz.	Messung, Marker ¹	Kameras ²	Szenarien bzw. unters. Parameter	Projekt, Kollab. ³
1	2005	Jülich, JSC	34	6	manuell	3, 5	Korridor	7
2	2005	Jülich, JSC	60	18	halbauto.	3, 5	Engstellenbreite	1, 7
3	2006	Düsseldorf, Berg. Kaserne	200	99	1	1, 3, 5	Korridor, Engstelle	1, 2, 7
4	2008	Kanpur, Indien	64	6	manuell	fremde	Kultur	3
5	2008	Wuppertal	50	14	halbauto.	6	Motivation	1
6	2009	Düsseldorf, Arena und Messehalle	350	170	2	3, 4, 6, 7	Tribünen, Treppen, Korridore, Ecken, Kreuzungen	1, 2, 8
7	2012	Wuppertal	33	10	3, 5	6, 7	Fahrradverkehr	4
8	2013	Tokio, Japan	28	260	3, 4	fremde	Wahlverhalten, Stauwellen, Kultur, Raumein-, austritt	5
9	2013	Düsseldorf, Messe	926	190	2	1, 2, 4, 6, 7	kritische Dichten	6, 9

Tab. 4.1: Durchgeführte Experimente; die Personenzahl bezieht sich auf die maximale Anzahl gleichzeitig an einem Versuch teilnehmender Personen; ¹Art der Messung: manuell, halbautomatisch oder vollautomatisch mittels Marker, Liste unterstützter Marker siehe Abb. 3.8 und 3.9; ²Liste benutzter Kameras siehe Tab. 4.2;

³Kooperationspartner:

- (1) Prof. Klingsch, Bergische Universität Wuppertal;
 - (2) Prof. Schadschneider, Universität zu Köln;
 - (3) Prof. Chakroborty, Indian Institute of Technology Kanpur, India;
 - (4) Prof. Huber, Bergische Universität Wuppertal;
 - (5) Prof. Nishinari, University of Tokio, Japan;
 - (6) Prof. Rusch, Universität Siegen;
- Projekte:
- (7) DFG KL 1873/1-1 und SE 1789/1-1 [525];
 - (8) Hermes [526, 527];
 - (9) BaSiGo [528]

werden. Dass dies sogar unter Laborbedingungen schwer zu erreichen ist kann z. B. [139] entnommen werden. Für eine statistisch aussagekräftige Anzahl an Messungen sind die Versuche mit möglichst vielen Personen durchzuführen.

Laborversuche wurden anstatt von Feldstudien durchgeführt, um

- gezielt den Einfluss einer die Dynamik beeinflussenden einzelnen Untersuchungsgröße (z. B. Personendichte, Engstellenbreite, Motivation) ändern und studieren sowie die restlichen Einflüsse (Probandenzusammensetzung, Lichtverhältnisse, Untergrund, Geräusche etc.) unverändert belassen zu können,
- die Möglichkeit der Parametervariation zu haben (z. B. Türbreite oder Personendichte beliebig einstellen und ein interessantes Parameterintervall kontinuierlich durchmessen zu können),
- die besonders sicherheitsrelevanten hohen Dichten, die in der Realität glücklicherweise selten

auftreten und somit selten beobachtet werden können, gezielt erzeugen zu können,

- die Probandenzusammensetzung durch Fragebögen (u. a. Alter, Fitness, Geschlecht, Kulturkreis, Größe, Zusammengehörigkeit, Psyche, Wohnort) beliebig genau erfassen zu können,
- in Raum und Zeit exakte Trajektorien durch die Nutzung von Markern und optimaler Lokalität (u. a. Ausleuchtung, Aufhängenhöhe) ermitteln zu können.

Die Versuchsaufbauten wurden meist künstlich hergestellt. Nur für die Hermes-Versuche (Nr. 6), die im Rahmen eines Projektes durchgeführt wurden, bei dem ein Evakuierungsassistent für ein bestimmtes Bauwerk entwickelt werden sollte [526, 527], wurden zudem Versuche im existierenden Bauwerk an für die Evakuierung relevanten und kritischen Gebäudeteilen durchgeführt (siehe Abb. 4.8 und 4.9), so dass eine Validierung der die Evakuierung beschreibenden mathematischen Modelle besonders genau durchgeführt werden konnte. Die untersuchten Gebäudeteile waren zudem schwer künstlich aufzubauen (Treppen) oder hatten eine spezielle, einmalige Geometrie (Tribüne mit Mundloch). Eine Liste der bisher durchgeführten Versuche kann Tab. 4.1 entnommen werden. Im Dokument wird bei Bezug auf einzelne Versuchsreihen neben der Nummer in Tab. 4.1 auch die Projektbezeichnung oder der Ort herangezogen: Versuche aus Versuchsreihe Nr. 3 werden als Kasernen-, Nr. 6 als Hermes-, Nr. 8 als Japan- und Nr. 9 als BaSiGo-Versuche bezeichnet.

Auch wenn die Laborversuche bestechende Vorteile haben, so dürfen die Nachteile nicht unerwähnt bleiben:

- der künstliche Aufbau und die Akquirierung der Probanden erfordern einen großen zeitlichen und finanziellen Aufwand,
- die Menge der die Dynamik beeinflussenden Parameter ist groß und kann nicht erschöpfend mit ggf. bestehenden Abhängigkeiten der Parameter untereinander untersucht werden
- bestimmte Parameter lassen sich schwer künstlich einstellen (z. B. Psyche, Physis der Personen), beispielsweise das Verhalten beim Verlassen eines Stadions nach einem klassischen Konzert von im Schnitt älteren das Stadion selten aufsuchenden Personen, oder regelmäßig wiederkommende Fans nach einem verlorenen Fußballspiel (ggf. unter Alkoholeinfluss).

Aus diesem Grund wurden ebenfalls Feldstudien (*in situ*) durchgeführt [532], die bisher manuell ausgewertet wurden und deren Daten sich auf reine Durchgangszeiten beschränken. Die markerlose Personendetektion, die in Abschnitt 3.5 vorgestellt wird, soll hierbei in Zukunft auch die Erfassung von Laufwegen in Feldstudien ermöglichen.

Da der Aufwand der Durchführung wohldokumentierter Versuche hoch ist und die erfassten Laufwege vielfältige Untersuchungen erlauben, sind die von uns analysierten Datensätze im Internet frei veröffentlicht worden [533]. Im laufenden BaSiGo-Projekt zur Versuchsreihe Nr. 9 ist geplant, alle erfassten Daten über eine Open-Access-Datenbank frei zugänglich zu machen.

4.1 Durchführung

4.1.1 Erfassungstechnik

Alle Versuche wurden senkrecht zur Bewegungsrichtung aufgezeichnet, egal ob von oben oder von der Seite (Ausnahmen bei Teilen der Versuchsreihe Nr. 6 und 7). Damit kann aufgrund des festen



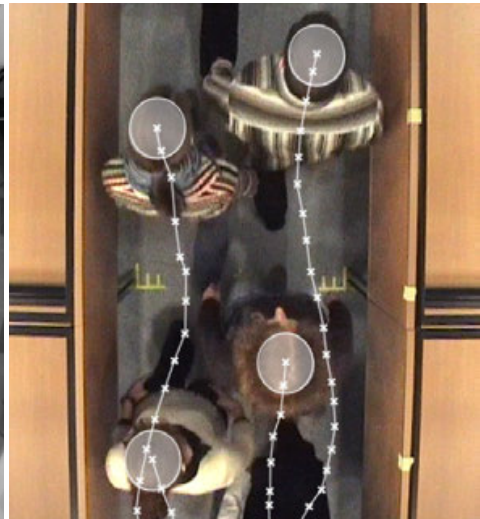
(a)



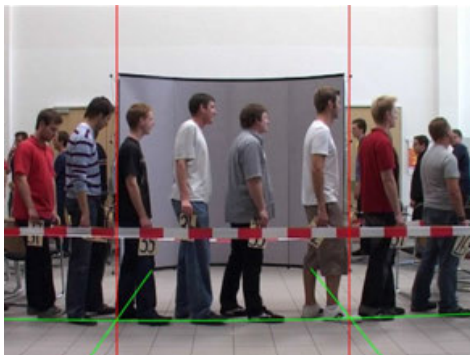
(b)



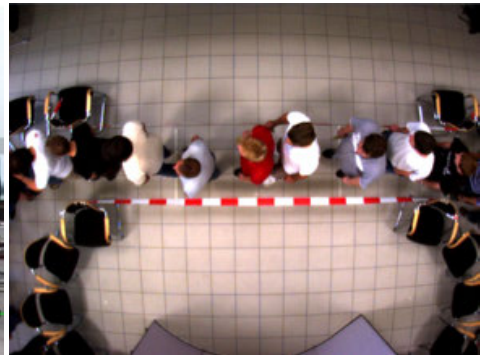
(c)



(d)



(e)



(f)

Abb. 4.1: Übersicht an Kameraeinstellungen der durchgeführten Versuche (zum Teil auf den Versuchsbereich beschnitten); Versuchsnummern gemäß Tab. 4.1: a: 1, b: 4, c, d: 2, e, f: 5; Videos zu entsprechenden Versuchen können über Tab. 4.3 erreicht werden; weitere Einstellungen in Abb. 4.2 und 4.3

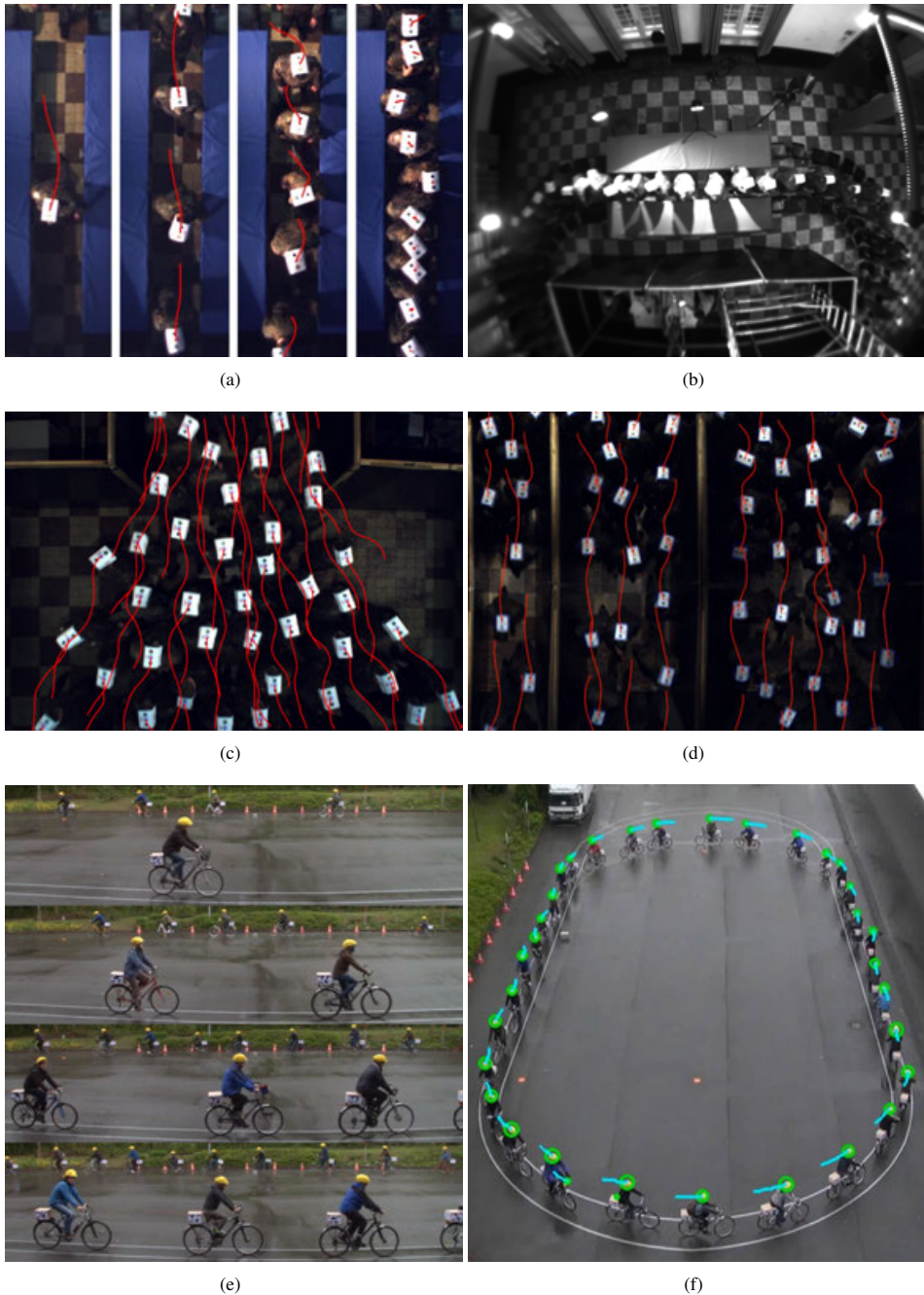


Abb. 4.2: Übersicht an Kameraeinstellungen der durchgeführten Versuche (zum Teil auf den Versuchsbereich beschnitten); Versuchsnummern gemäß Tab. 4.1: a-d: 3, e, f: 7; Videos zu entsprechenden Versuchen können über Tab. 4.3 erreicht werden; weitere Einstellungen in Abb. 4.1 und 4.3

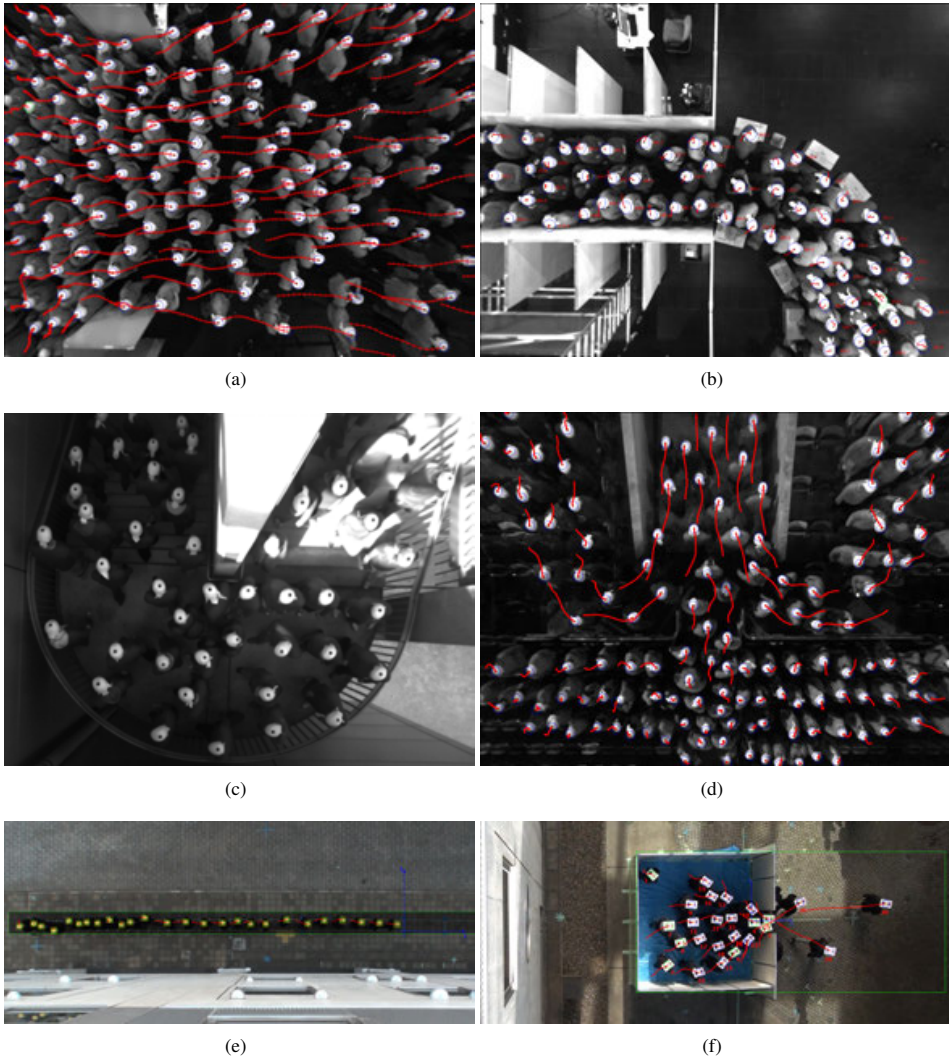


Abb. 4.3: Übersicht an Kameraeinstellungen der durchgeführten Versuche (zum Teil auf den Versuchsbereich beschnitten); Versuchsnummern gemäß Tab. 4.1: a-d: 6, e, f: 8; Videos zu entsprechenden Versuchen können über Tab. 4.3 erreicht werden; weitere Einstellungen in Abb. 4.1 und 4.2

Kameraabstandes die Position einer Person mit bekannter Körpergröße direkt aus der Position im Bild bestimmt werden. Die genaue Berechnung ist in Abschnitt 3.1 und 3.4 nachzulesen.

Bei der Seitenansicht der eindimensionalen Versuche sind insbesondere für die manuelle Erfassung reiner Ein- und Austrittszeiten entweder real oder nachträglich in einem Videoschnittprogramm Stäbe bzw. Hilfslinien aufgestellt bzw. eingezeichnet worden. Dabei ist wichtig, dass sich die Personen im gleichen Abstand zur Kamera bewegen [535–537].

4.1.1.1 Verdeckung

Bei den nicht eindimensionalen Versuchen ist die Betrachtung von oben zwingend, um alle Personen zu jeder Zeit im Sichtfeld der Kamera sehen zu können. Wegen der perspektivischen Sicht einer Kamera steigt jedoch aufgrund von Größenunterschieden bei den Personen zum Rand die Rate der Verdeckung. Durch eine orthogonale Kameraperspektive ist die Wahrscheinlichkeit einer Verdeckung am geringsten, da bei schräger Sicht der Einfallswinkel flacher wird und die Köpfe sich eher voneinander schieben. Da wir bei Laborversuchen die Perspektive selber bestimmen können, sind alle Aufnahmen genau senkrecht durchgeführt worden. Für die orthogonale Ausrichtung ist das Lot von der Kamera gefällt worden und der Fußpunkt durch Neigung der Kamera in die Bildmitte verschoben worden. Ausgenommen der Versuche an Treppen ist darauf geachtet worden, dass der Boden waagerecht ist, nicht nur für eine genaue Lokalisierung der Personen, sondern auch, um dadurch die Dynamik nicht zu beeinflussen.

Für die Betrachtung der Verdeckung gehen wir von einer maximalen Personendichte von ρ aus. Bei einer optimalen Gleichverteilung von $\sqrt{\rho} \cdot \sqrt{\rho}$ Personen in der Fläche ergibt sich ein Abstand zwischen den Personen von

$$d_p = \frac{1}{\sqrt{\rho}}. \quad (4.1)$$

Eine reale Verteilung der Personenabstände eines Engstellenversuchs ist Abb. 3.104 zu entnehmen.

Die durchschnittliche Kopfbreite und -länge eines Deutschen beträgt heutzutage ca. $15 \text{ cm} \times 20 \text{ cm}$ (siehe Abb. 3.17 in [492]). Die winkelabhängige Ausdehnung der Verdeckung des ellipsoidischen Kopfes werde vereinfacht mit dem festen Wert $w_k = 15 \text{ cm}$ angenommen. Zur Erkennung eines Markers muss ein Bereich von $w_m \times w_m$ einzusehen sein.

Für einen Winkel α von der optischen Achse des Aufnahmesystems ergibt sich daraus aufgrund der perspektivischen Abbildung ein maximal erlaubter Größenunterschied d_h (siehe Abb. 4.4) von

$$d_h = \frac{d_p - w_m/2 - w_k/2}{\tan \alpha}. \quad (4.2)$$

Mit einem vorgegebenen Größenunterschied der Probanden von d_h ist damit der maximal einzusehende Öffnungswinkel bzw. Bildwinkel 2α

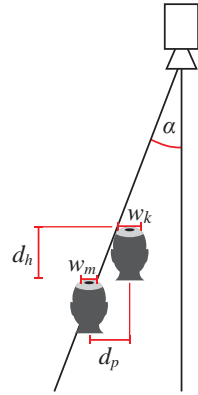


Abb. 4.4: Maximaler Bildwinkel 2α ohne Verdeckung bei einem minimalen Personenabstand von d_p und einer maximalen Größendifferenz von d_h

$$2\alpha = 2 \arctan \frac{2d_p - w_m - w_k}{2d_h}. \quad (4.3)$$

Daraus ergibt sich beispielsweise für die ebenen Hermes-Versuche (Nr. 6) mit einer maximalen Dichte von $\rho = 4/\text{m}^2$ und damit einem Personenabstand von $d_p = 50\text{ cm}$, einem laut ausgehändigten Fragebögen maximalen Größenunterschied von $d_h = 60\text{ cm}$ (siehe auch Abschnitt 3.6.2) und einem für den eingesetzten Hermes-Marker (Nr. 2) (siehe Abb. 3.8) einzusehenden Bereich von $w_m \times w_m = 10\text{ cm} \times 10\text{ cm}$ ein Bildwinkel, bei dem alle Marker gut sichtbar sind, von

$$2\alpha = 2 \arctan \frac{75}{120} \approx 64^\circ.$$

Dies entspricht ungefähr dem horizontalen Bildwinkel der eingesetzten Kamera Nr. 4 (siehe Tab. 4.2).

Trotzdem kann es zu Verdeckungen sogar bei einem geringeren Winkel kommen, da Personen nicht gleichverteilt laufen und der Marker durch schiefes Tragen oder Neigen des Kopfes nicht in der Kopfmittle liegt.

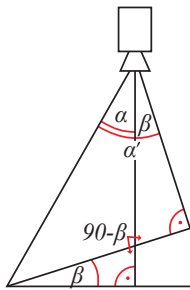


Abb. 4.5: Einschränkung des Bildwinkels bei senkrechten Aufnahmen an Treppen

Der einzuhaltende Abstand der Personen zueinander ist für den einfachen robusten Farbmarker größer, da die dabei für gewöhnlich benutzten farbigen Mützen im Bild komplett voneinander getrennt sein müssen. Mit den Randbedingungen der ebenen Hermes-Versuche (Nr. 6) und einem aufgrund des Farbmarkers erweiterten einzusehenden Bereiches von ca. $w_m \times w_m = 30\text{ cm} \times 30\text{ cm}$ ergibt sich ein einsehbarer Bildwinkel von $2\alpha \approx 49^\circ$.

Eine weitere Einschränkung des Bildwinkels ergibt sich bei Versuchen an Treppen (Teile der Versuchsreihe Nr. 6). Bei einem Steigungswinkel der Treppe von β erhöht sich der theoretische Bildwinkel α' auf $\alpha + \beta$ (siehe Abb. 4.5) bzw. muss der eigentliche Bildwinkel für eine möglichst verdeckungslose Aufzeichnung um den Steigungswinkel eingeschränkt werden. Daher wurde für Versuche an Treppen ein Überlappungsbereich der Kameras entsprechend groß gewählt (siehe Abb. 4.8). Zudem ist durch das Treppensteigen der Kopf der Person oft geneigt und damit der Marker im unteren Teil der Treppe schlechter sichtbar. Es hat sich zudem gezeigt, dass die Anordnung der Treppenstufen den

Personenabstand dominiert und aufgrund der Tatsache, dass das Gesicht durch leichten Versatz freie Sicht hat, der Abstand beim Treppenabstieg in Laufrichtung der Stufenlänge entspricht (siehe Video Nr. 39 und 22). Trotz der häufig leichten Kopfneigung wurde die Kamera auch bei den Versuchen an Treppen so aufgehängt, dass die optische Achse der Kamera zum Lot ausgerichtet war, da dadurch die Marker auf dem Kopf am besten gesehen werden konnten.

4.1.1.2 Sichtfeld und überlappende Kameras

Für ebene Versuche ist nach Festlegung des Bildwinkels und mit der Aufhängehöhe der Kamera die in Kopfhöhe durch ein Aufnahmesystem abdeckbare rechteckige Fläche einfach zu bestimmen. Beispielsweise wurden für die Planung der BaSiGo-Versuche (Nr. 9) Objektive mit einem horizontalen Bildwinkel (*horizontal angle of view* bzw. häufiger fälschlich benutzt *horizontal field of view*, HFOV) von 32° ausgewählt, da eine sehr hohe maximale Dichte von $\rho = 10/\text{m}^2$ angestrebt wird. Die Dichte führt laut Gleichung 4.1 zu einem Personenabstand von $d_p \approx 32\text{ cm}$. Bei der Akquise der Probanden wird darauf geachtet, dass die Personengröße zwischen 145 cm und 205 cm liegt; somit ist $d_h = 60\text{ cm}$. Der Codemarker wird einen einzusehenden Bereich von $w_m = 15\text{ cm}$ brauchen. Zusammen ergibt sich (siehe Kamera Nr. 2 in Tab. 4.2) ein Bildwinkel von

$$2\alpha = 2 \arctan \frac{34}{120} \approx 32^\circ.$$

Bei einer Aufhängehöhe der Kamera von 750 cm ist die Sichtfeldbreite in einer Entfernung der durchschnittlichen Personengröße von 176 cm

$$2 \cdot (750 - 176) \text{ cm} \cdot \tan 16^\circ \approx 329 \text{ cm}.$$

Bei einem Seitenverhältnis von 5:4 des Kameraschips ergibt sich damit eine abdeckbare rechteckige Fläche von $l_h \times l_v = 329 \text{ cm} \times 263 \text{ cm}$.

Ein Versuchsaufbau der BaSiGo-Versuche (Nr. 9) hat eine Grundfläche von $l_g \times l_g = 10 \text{ m} \times 10 \text{ m}$. Demnach sind überlappende Kameras nötig, um das gesamte Versuchsfeld zu erfassen. Für eine saubere Verbindung der Laufwege der Kameras mit überlagertem Sichtfeld ist ein Überlappungsbereich von $l_u \approx 100 \text{ cm}$ sinnvoll, wobei in der horizontalen Achse eher eine größere und in der vertikalen Achse eine geringere Überlagerung nötig ist (hier: 107 cm bzw. 80 cm). Die Anzahl der Kameras n_k in eine Richtung ergibt sich aus

$$l_{h,v} + (l_{h,v} - l_u)(n_k - 1) = l_g. \quad (4.4)$$

Horizontal sind damit aufgerundet 4 und vertikal 5 Kameras nötig.

Damit die Fehler bei der Positionsbestimmung

- aufgrund der perspektivischen Sicht,
- aufgrund der optischen Verzerrung und
- aufgrund von Verdeckungen,

die alle von der optischen Achse zum Rand hin zunehmen, möglichst klein sind, wurden die Kameras in den Zeilen auf Lücke angeordnet (siehe Abb. 4.6). Zwei Kameras müssen dafür zur Abdeckung der Grundfläche hinzugenommen werden, dienen hier aber gleichzeitig der besseren Sicht auf die dort positionierten Ein- und Ausgänge.

Die Punkte, die am weitesten von den umliegenden optischen Achsen der Kameras entfernt liegen, besitzen dann den geringsten Abstand zu diesen Achsen, wenn die Kameras versetzt gemäß Abb. 4.7 b angeordnet sind. Dies ist geometrisch leicht einzusehen, wenn man den roten

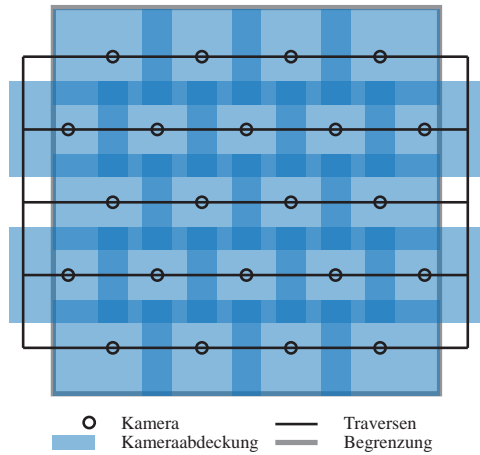


Abb. 4.6: Kamera-Grid zur Abdeckung einer Versuchsfläche von $10 \text{ m} \times 10 \text{ m}$

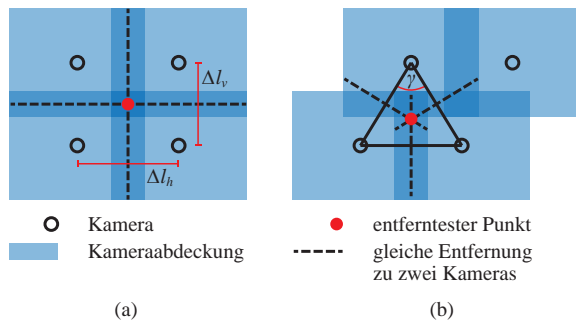


Abb. 4.7: Verringerung des maximalen Winkels zur optischen Achse durch Änderung der Anordnung der überlappenden Kameras von einer (a) rechteckigen zu einer (b) versetzten Aufhängung

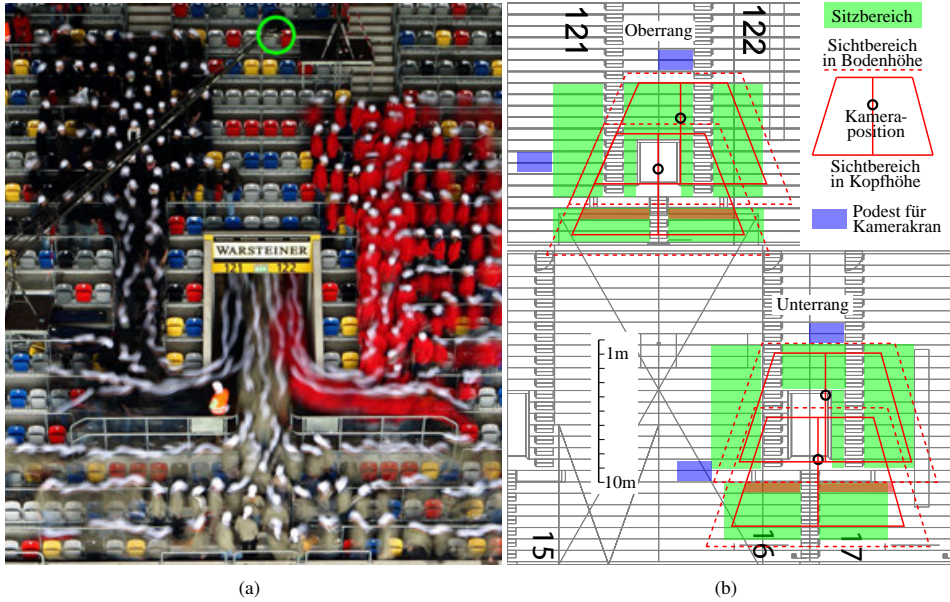


Abb. 4.8: Experimente auf der Tribüne der ESPRIT Arena: (a) 300 Personen verlassen die Tribüne im Oberrang durch ein Mundloch; der grüne Kreis zeigt die Position einer am Kamerakran angebrachten Stereokamera (Langzeitbelichtung, Quelle: Jan Ovelgönne); die Farbe der Oberhemden diente der Unterscheidung, aus welchem Block die Probanden kamen; siehe auch Video Nr. 39 (b) Skizze der Aufbauten der Erfassungstechnik (Aufsicht), die die Position der zwei Stereokameras (schwarze Kreise), die Aufnahmebereiche in Kopf- (rote durchgehende Linie) und Bodenhöhe (rote gestrichelte Linie) sowie den Aufstellungsort der Kamerakräne (blaue Fläche) zeigt

Punkt, der den Punkt mit dem weitesten Abstand zu den Kameras darstellt, in Abb. 4.7 a und b betrachtet. Dieser verschiebt sich zwischen a und b nach unten, wodurch der Abstand zum Bildmittelpunkt der Kameras kleiner wird.

Der Abstand der Kameras in den Zeilen und Spalten des Kamera-Grids ist $\Delta l_{h,v} = l_{h,v} - l_u$. Bei einer rechteckigen Anordnung gemäß a ist der entfernteste Punkt die Mitte der die Kameras verbindenden Diagonalen und damit der Abstand zur Bildmitte

$$r_{\square} = \frac{1}{2} \sqrt{\Delta l_h^2 + \Delta l_v^2}. \quad (4.5)$$

Bei der verwendeten Anordnung b ist der von allen Kameras am weitesten entfernt liegende Punkt innerhalb der jeweils durch die versetzten Kameras aufgespannten Dreiecke der Schnittpunkt der Mittelsenkrechten dieser Dreiecke, wenn dieser innerhalb des Dreiecks liegt. Der Schnittpunkt ist der Umkreismittelpunkt, dessen Radius durch den Sinussatz gegeben ist:

$$r_{\triangle} = \frac{\Delta l_h}{2 \sin \gamma}. \quad (4.6)$$

Sollte der Umkreismittelpunkt nicht innerhalb des Dreiecks liegen ($\gamma > 90^\circ$), so gibt es zwei Punkte, die am weitesten von der optischen Achse der Kameras entfernt liegen. Diese sind die Schnittpunkte

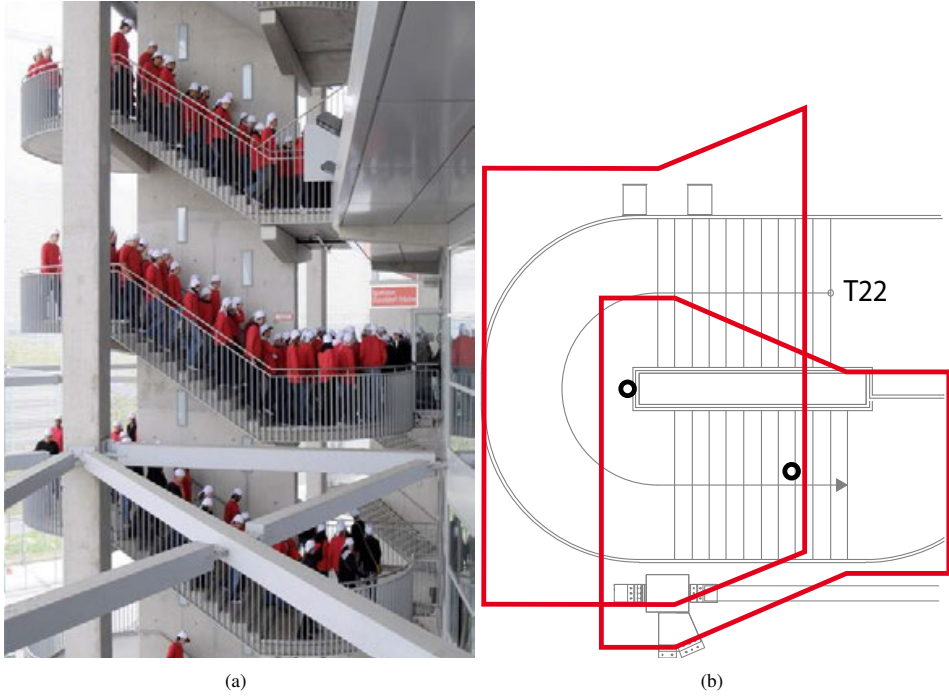


Abb. 4.9: Experimente an der Außentreppe der ESPRIT Arena: (a) beim Verlassen im Evakuierungsfall; (b) Skizze der Abdeckung der Erfassungstechnik in Kopfhöhe (Aufsicht); Symbole entsprechend Abb. 4.8

der Mittelsenkrechten der kurzen Dreiecksseiten mit der langen Dreiecksseite. Da dafür gelten muss, dass

$$\tan 45^\circ < \frac{\Delta l_h}{2\Delta l_v} \Leftrightarrow \Delta l_h > 2\Delta l_v,$$

kann dieser Fall vernachlässigt werden, weil gängige Seitenverhältnisse von Sensoren auch nach Abzug der Überlappungsbereiche das Seitenverhältnis von 2:1 nicht überschreiten.

Für den Grenzfall $\gamma = 90^\circ \Leftrightarrow \Delta l_h = 2\Delta l_v$, gegen die der Abstand kontinuierlich strebt, kann gezeigt werden, dass

$$\begin{aligned} r_{\square} &> r_{\Delta} \\ \Leftrightarrow \frac{1}{2} \sqrt{\Delta l_h^2 + \Delta l_v^2} &> \frac{\Delta l_h}{2 \sin \gamma} \\ \Rightarrow \frac{\sqrt{5}}{4} \Delta l_h &> \frac{1}{2} \Delta l_h. \end{aligned}$$

Der Abstand verringert sich somit um den Faktor von $2/\sqrt{5}$.

An reinen Schrägen bzw. Treppen ergibt sich, wie exemplarisch in Abb. 4.8 zu sehen, eine trapezförmige Abdeckung. Dabei zeigt die gestrichelte rote Linie die Abdeckung auf dem Boden, die durchgezogene Linie die entscheidende Abdeckung in Kopfhöhe. Für andere Untergründe ist die Form beliebig (siehe Abb. 4.9).

4.1.1.3 Brennweite und Sensorgröße

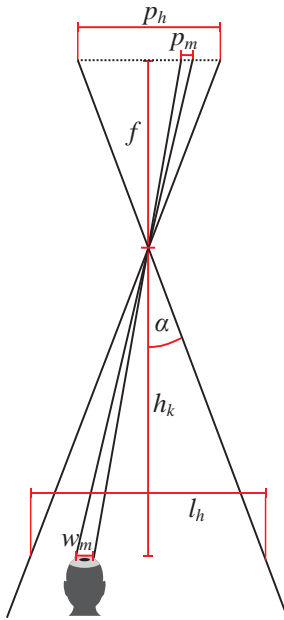


Abb. 4.10: Bestimmung der Pixelabdeckung $p_m \times p_m$ und Brennweite f bei bekannter äußerer Geometrie α , h_k und w_m auf Basis des Lochkammermodells

Ist der benötigte horizontale Bildwinkel 2α und der maximale Kopf-
abstand zur Kamera h_k und damit $l_h = 2h_k \tan \alpha$ bekannt, so kann bei
einer benötigten Pixelabdeckung der Markerfläche von $p_m \times p_m$ (qua-
dratisch, da Marker beliebig im Bild liegen kann; siehe auch Spalte b
in Abb. 3.8) die bei verwendeter Optik erforderliche sichtbare hori-
zontale Sensoraufösung p_h und die Brennweite f der Optik bestimmt
werden (siehe Abb. 4.10):

$$\frac{p_h}{l_h} = \frac{p_m}{w_m} \Leftrightarrow p_h = l_h \frac{p_m}{w_m}, \quad (4.7)$$

$$\frac{p_h}{l_h} = \frac{f}{h_k} \Leftrightarrow f = h_k \frac{p_h}{l_h}. \quad (4.8)$$

Die vertikalen Werte ergeben sich entsprechend. Die so bestimmte
Brennweite hat die Einheit Pixel. Für ein sensorunabhängiges Ob-
jektiv wird die Brennweite jedoch metrisch angegeben. Zur Bestim-
mung der metrischen Brennweite ist zum einen die Sensorgröße bzw.
die Pixelgröße nötig und zum anderen die relative Abdeckung des
Sensors vom Bildkreis des Objektivs ohne Vignettierung.

Die Abdeckung bzw. nutzbare Sensorgröße ist immer geringer als die
angegebene Sensorgröße und beispielsweise für die Kamera Nr. 1
aus Tab. 4.2 mit der angegebenen Sensorgröße von $1/3'' \approx 8,5 \text{ mm}$
nur $p_h \times p_v = 4,8 \text{ mm} \times 3,6 \text{ mm}$ groß [538]. Mit dem verwendeten
Pentax-Objektiv mit einer Brennweite von $f = 4,2 \text{ mm}$ ergibt sich
daraus z. B. ein horizontaler Öffnungs- bzw. Bildwinkel von

$$2\alpha = 2 \arctan \frac{p_h}{2f} = 2 \arctan \frac{4,8 \text{ mm}}{2 \cdot 4,2 \text{ mm}} \approx 59^\circ. \quad (4.9)$$

Eine große Sensorgröße hat den Vorteil, dass der Sensor lichtempfindlicher ist und damit für gewöhn-
lich eine genauere Abbildung erlaubt, jedoch führt dies auch dazu, dass bei großer Blende die Tie-
fenschärfe geringer ausfällt und damit der Schärfebereich genau auf die Personengröße abgestimmt
werden muss.

4.1.1.4 Belichtung und Beleuchtung

Bei der Aufnahme muss darauf geachtet werden, dass die Lichtmenge über die Belichtungszeit bzw.
Verschlusszeit und Blende so eingestellt wird, dass die zu detektierenden Objekte scharf und kontrast-
reich abgebildet werden. Die Tiefenschärfe muss über die Blende so eingestellt werden, dass insbe-
sondere die Köpfe der unterschiedlich großen Personen scharf abgebildet werden. Die Belichtungszeit
soll zwar so groß sein, dass die Objekte kontrastreich erscheinen, jedoch soll die Bewegungsunschär-
fe für die zu erwartende Geschwindigkeit gering gehalten werden. Hier muss ggf. ein Kompromiss
zwischen Kontrast, Bewegungs- und Tiefenschärfe gefunden werden. Abb. 4.11 zeigt die Bewe-
gungsunschärfe bei einem durch das Bild rennenden Probanden bei den Kasernen-Versuchen (Nr. 3).

Um die Belichtungszeit für ein schärferes Bild verringern zu können ist eine gute Ausleuchtung sinnvoll. Zudem kann eine homogene Ausleuchtung dazu beitragen, die Robustheit der Personenerkennung durch Homogenisierung der Lichtverhältnisse zu erhöhen. Die Ausleuchtung sollte diffus bzw. indirekt sein und gleichmäßig von allen Seiten erfolgen. Dies ist am stabilsten mit rein künstlichem Licht realisierbar. Optimal ist jedoch ein gleichbleibend bewölkter Himmel.

Dass dies schwer zu realisieren ist, sieht man an den Aufnahmen in Abb. 3.8. Für Marker 1 ist die Aufnahme im Zentrum heller als am Rand. Die Beleuchtung erfolgte zudem vornehmlich durch Strahler von oben, so dass geneigte Köpfe einen wesentlich kontrastärmeren Marker aufweisen. Die Belichtung wurde so eingestellt, dass der Marker möglichst gut zu erkennen ist, was dazu führt, dass die Umgebung sehr dunkel ausfällt. Der abgebildete Versuch mit Marker 2 wurde im Stadion bei geöffnetem Tribünendach durchgeführt. Die starke Sonneneinstrahlung führte dazu, dass die Köpfe von einer Seite wesentlich stärker beleuchtet wurden. Ebenso ist in Abb. 4.3 c zu beobachten, dass Experimente im Freien bei starkem Sonnenschein zu großen Helligkeitsunterschieden bis hin zur Überbelichtung führen können. Auch bei Marker 3 ist in Abb. 3.8 die Ausleuchtung inhomogen, da ein Teil der Versuchsfläche direkt von der Sonne beschienen wurde.

Spielt die Farbe des Markers eine wichtige Rolle (Kasernen-Marker Nr. 1), so muss darauf geachtet werden, dass entweder die Beleuchtung über den ganzen Versuchszeitraum konstant gehalten wird oder aber bei Änderung ein Weißabgleich durchgeführt wird. Insbesondere die Lichtart und damit die Farbtemperatur (z. B. Sonne, Glühlampe, Leuchtstoffröhre) verändert ansonsten die Farbe zwischen den einzelnen Versuchen einer Versuchsreihe.



Abb. 4.11: Bewegungsunschärfe bei rennender Person, wenn die eingestellte Belichtungszeit eine Markerkennung nur von gehenden Personen erlaubt

4.1.1.5 Bildwiederholrate

Die letzte bei der Kameraauswahl zu berücksichtigende Größe ist die Bildwiederholrate. Die Anzahl der Einzelbilder pro Sekunde [f/s] (*frames per second*, auch *fps*) sollte so groß gewählt sein, dass

- der Weg einer Person zwischen den Einzelbildern als linear angenommen werden kann, um möglichst exakte Trajektorien zu erfassen,
- das automatische Verfolgen robust abläuft, was auch damit zusammenhängt, dass
- die im Folgebild erkannten Personen ihren vorherigen Positionen im Bild näher sind als andere Personen.

Alle Versuche aus Tab. 4.1 wurden mit Bildwiederholraten von 15, 16, 25 (PAL) oder 29,97 (NTSC) Bildern pro Sekunde (siehe auch Tab. 4.2) durchgeführt. Bei einer freien Gehgeschwindigkeit von 1,5 m/s ist eine Person somit zwischen zwei aufeinanderfolgenden Bildern maximal

$$\frac{1,5 \text{ m/s}}{15 \text{ f/s}} = 0,1 \text{ m/f}$$

weit gegangen, so dass immer die zwei am nächsten zueinander liegenden erkannten Personen in aufeinanderfolgenden Bildern dieselbe Person sind und somit zu einer Trajektorie gehören. Zudem

Nr.	Hersteller	Typ	Anschluss o. Medium	Sensorgröße (Farbe)	Auflösung (par)	f/s	Optik, Brennw.
1	IDS	UI-2230-C [539]	USB 2.0	1/3" CCD (Color, Bayer)	1024x768 (1:1)	25	Pentax, 4,2 mm Tamron, 4,0 mm
2	IDS	UI-3240CP-M [540]	USB 3.0	1/1,8" CMOS (Mono)	1280x1024 (1:1)	25	Pentax, 12,0 mm
3	Point Grey	Bumblebee [541]	IEEE-1394a	2x 1/3" CCD (Mono)	1024x768 (1:1)	15	2,0 mm
4	Point Grey	BBX3-13S2M-38 [542]	IEEE-1394b	3x 1/3" CCD (Mono)	1280x960 (1:1)	16	3,8 mm
5	Sony	DCR-VX2000E [543]	DV-Band	1/3" CCD (Color, 3C)	704x576 (12:11)	25	6,0 mm
6	Sony	HVR-Z5E [544]	IEEE-1394a	1/3" CMOS (Color, 3C)	1440x1080 (4:3)	25	4,1 mm
7	Sony	CX740VE [545]	SD	1/2,88" CMOS (Color)	1440x1080 (4:3)	25	3,8 mm

Tab. 4.2: Eingesetzte Kameras und vorgenommene Einstellungen; die Auflösung ist die für die Aufzeichnung verwandte und bei Industriekameras (1-4) die native Sensorauflösung, bei Camcordern (5-7) ist dies die maximale Auflösung für Bewegtbilder; par (*pixel aspect ratio*) ist das Pixelseitenverhältnis; f/s oder fps (*frames per second*) bezeichnet die Bildwiederholrate, die zur Aufzeichnung verwandt wurde und ist bis auf Nr. 2 die maximale Rate; die Optik ist entweder fest verbaut oder ein Objektiv über C-Mount-Fassung; die Brennweite ist zum Teil variabel, dabei wurde immer die angegebene minimale Brennweite eingesetzt; die Brennweite im Verhältnis zur Sensorgröße gibt Aufschluss über den Bildwinkel, wobei die nutzbare Sensorgröße kleiner ist [538]

erlaubt eine Wiederholrate von mindestens 15 f/s ein für das menschliche Auge flüssiges Bild. Darunter würden die Einzelbilder als solche wahrgenommen werden.

Die benötigte Bildwiederholrate und die maximale Datenrate der verwendeten Übertragungsschnittstelle begrenzen die maximal mögliche Sensorgröße.

4.1.1.6 Sonstige Kameraeigenschaften

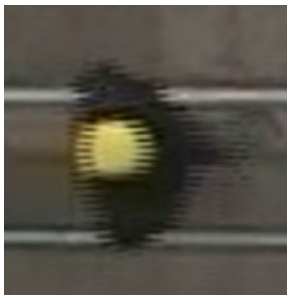


Abb. 4.12: Ausgefranzte Kopfform durch die Aufzeichnung im Zeilensprungverfahren

Alle verwandten Industriekameras nehmen das einzelne Bild immer fortlaufend (*progressive*) auf, bei den DV- und HDV-Kameras ist dies ebenfalls eingestellt worden. Dies dient dazu, dass aufeinanderfolgende Zeilen keinen zeitlichen Versatz haben und somit Formen in der Bewegung besser erkannt werden können. Bei Aufnahmen mit fremden Kameras (Versuchsreihe Nr. 4 und 8) war dies nicht immer möglich. Abb. 4.12 zeigt solch einen Versuch, bei dem die Aufnahme im Zeilensprungverfahren (*interlaced scanning*) aufgenommen werden musste und dadurch die Bewegung die Form des Kopfes ausfranst. Eine Interpolation jeder zweiten Zeile (*deinterlacing*) verbessert die Form, führt aber zu einem Schärfeverlust.

Farbsensoren mit Bayer-Filter sind nur genommen worden, wenn die Farbinformation wichtig war (Kamera Nr. 1, Versuch Nr. 3, Marker Nr. 1, siehe Abschnitt 3.4.1), ansonsten liefern monochrome Sensoren ein helleres und schärferes Bild. Nur die professionelle DV- bzw. HDV-Kamera Nr. 5 und 6 haben ein 3-Chip-System, das jeden Farbkanal über einen separaten Sensor erfasst.

Vor der Analyse muss bei den Camcordern das Pixelseitenverhältnis auf quadratische Pixel umgerechnet werden, damit im Bild ein einheitliches Pixel-Meter-Verhältnis herrscht. Bei Kamera Nr. 5 ist

eine horizontale Pixelstreckung (vertikale Stauchung wurde nicht vorgenommen, um keine Bildinformationen zu verlieren) um einen Faktor von 12/11 und bei Kamera Nr. 6 und 7 um einen Faktor von 4/3 nötig. Dies kann außerhalb der in dieser Arbeit vorgestellten Software *PeTrack* geschehen oder innerhalb *PeTrack* über sich unterscheidende Brennweiten in horizontaler und vertikaler Richtung realisiert werden (siehe Abschnitt 3.1.1).

4.1.1.7 Übertragungsschnittstelle

Nachdem die kameraeigenen Eigenschaften besprochen wurden, soll nun die zur Erfassung benötigte Peripherie erörtert werden.

Die Industriekameras (Nr. 1 bis 4) besitzen keine integrierte Speichereinheit, so dass für eine Aufzeichnung die Bilder im Normalfall an einen Rechner übertragen werden müssen. Die Camcorder hingegen besitzen Aufzeichnungsmöglichkeiten direkt an der Kamera: als Medium dient der Kamera Nr. 5 ein DV-Band, Nr. 6 eine über eine IEEE-1394a-Schnittstelle angeschlossene Festplatte und Nr. 7 eine SD-Speicherkarte. Für Industriekameras sind die folgenden Übertragungsschnittstellen verbreitet: analog, Camera Link, USB 2.0, USB 3.0, IEEE 1394a (FireWire 400), IEEE 1394b (FireWire 800), GigE Vision, SD-SDI und HD-SDI. Jede Schnittstelle hat ihre Vor- und Nachteile, dabei können diese aufgeschlüsselt werden in [546]:

Standardisierung: im Gegensatz zu proprietären Lösungen ist eine standardisierte Schnittstelle für gewöhnlich zukunftssicherer und leichter erweiterbar,

Verfügbarkeit: ist die Anzahl der verfügbaren Kameras mit dieser Schnittstelle groß, so können am ehesten die oben erarbeiteten Kameraparameter realisiert werden,

Kosten: der Gesamtpreis aller Komponenten einschließlich Kamera, Kabel, Bilderfassungskarte im Rechner und Stromverbrauch,

Durchsatz: die Netto-Geschwindigkeit (Abzug von Protokoll-Zuschlag), mit der Bilddaten über die Schnittstelle übertragen werden können; sichergestellte Datenrate,

Kabellänge: die maximal mögliche Kabellänge mit und ohne zusätzlichen Verstärkern,

Stromversorgung: die Stromversorgung kann über das Datenkabel bereitgestellt werden oder muss separat erfolgen,

Synchronisation: einige Bussysteme unterstützen die Synchronisation angeschlossener Kameras,

Prozessorauslastung: Auslastung des Prozessors des Rechners bei der Bilderfassung.

Die vom Autor ausgewählten Schnittstellen können der Tab. 4.2 entnommen werden. IEEE-1394 wurde aufgrund des Punktes „Verfügbarkeit“ genommen, da die Stereokamerasysteme nur mit dieser Schnittstelle ausgeliefert wurden. USB besitzt große Vorteile im Bereich der „Standardisierung“, „Verfügbarkeit“, „Kosten“, „Stromversorgung“ und mit USB 3.0 auch im Bereich des „Durchsatzes“. Die größten Probleme gab es bei beiden verwandten Schnittstellen mit der „Kabellänge“, da diese ohne Verstärker (*repeater*) oder Verwendung von speziell konfektionierten oder aktiven Kabeln nur rund 5 m reichten. Daher wurde in allen Fällen mit aktiven Verstärkern gearbeitet, um die benötigten Kabellängen von über 10 m zu erreichen (siehe auch Abb. 4.14). USB erzeugt zudem im Verhältnis zu den anderen Schnittstellen eine hohe Prozessorlast, was auch bei den Kasernen-Versuchen (Nr. 3) dazu geführt hat, dass es zu Fehlbildern kam (siehe Abschnitt 3.3.3).

4.1.1.8 Synchronisation

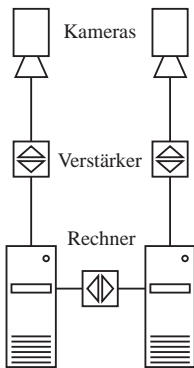


Abb. 4.13: Kabelwege der Stereokameras in Versuchsserie Nr. 6 zur Bildübertragung und Synchronisation

Bei der Erfassung eines zusammenhängenden Gebietes mit mehreren Kameras ist eine Synchronisation für eine möglichst genaue Verbindung überlappender Laufwege in Ort und Zeit sinnvoll (siehe Abschnitt 3.3.5). Für die Hermes-Versuche (Nr. 6) wurden softwaretechnisch unter Zuhilfenahme des FireWire-Busses IEEE1394b die Stereokameras Nr. 4 synchronisiert [547]. Hierzu wurden die aufnehmenden PC-Systeme mit einem weiteren FireWire-Kabel verbunden und beim Herausschreiben der Bilddaten parallel der Zeitstempel eines jeden Bildes gespeichert (siehe Abb. 4.13). Bei den BaSiGo-Versuchen (Nr. 9) werden die Kameras Nr. 2 über ein separates Netz über deren GPIO-Anschluss synchronisiert.

Ist eine direkte Synchronisation technisch nicht realisierbar (z. B. bei Kameras Nr. 5, 6 und 7), so kann unter gewissen Voraussetzungen gemäß Abschnitt 3.3.5 eine automatische örtliche wie zeitliche Abstandsminimierung zwischen den Trajektorien der unterschiedlichen Kameras durchgeführt und damit der Versatz bestimmt werden. Als Voraussetzung für ein verlässliches Ergebnis darf der örtliche und zeitliche Versatz nicht zu groß und die Anzahl der gleichzeitig sichtbaren detektierten Personen nicht zu klein sein. Sind die Voraussetzungen nicht erfüllt, so kann der zeitliche Abgleich manuell stattfinden. Dabei ist es sinnvoll, mit allen zu synchronisierenden Kameras einen einmaligen schnellen

Vorgang aufzuzeichnen (z. B. Hände klatschen, Licht an- oder ausschalten), an der die Bildsequenzen zeitlich gegeneinander ausgerichtet werden können. Kann bei den eingesetzten Kameras der Ton mit aufgezeichnet werden, so ist aufgrund der nutzbaren Abtastrate des Tons (z. B. 44,1 kHz) gegenüber der nutzbaren Bildwiederholrate vom Video (z. B. 25 Hz) ein akustisches Signal vorzuziehen.

4.1.1.9 Speicherung

Alle Versuche wurden zur späteren Analyse aufgezeichnet. Da die inhaltliche Auswertung unabhängig von der Bildanalyse auch an den archivierten Bilddaten erfolgen sollte, ist die automatische Erfassung der Laufwege nicht auf Echtzeitverfahren angewiesen. Zudem ist dadurch eine Sichtung der Ergebnisse der Verfahren sowie eine ggf. nötige Korrektur möglich, so dass die Ausgangsdaten für die inhaltliche Analyse ohne Fehler sind.

Bei den Camcordern ist die Kodierung und Speicherung direkt an den Aufnahmesystemen gegeben, für die Industriekameras müssen die Bilddaten von der Kamera bis zum Speicher transportiert und ggf. weiterverarbeitet werden.

Auch wenn die Camcorder den Vorteil der direkten rechnerunabhängigen Aufzeichnung aufweisen, hat der Einsatz von Industriekameras wesentliche Vorteile:

- kleine Bauform,
- Zusammenstellung von Sensor und Optik genau nach Bedürfnissen möglich,
- oftmals Möglichkeiten der Synchronisation vorgesehen.

Alle Komponenten in der Transportkette müssen gut aufeinander abgestimmt sein und Engstellen frühzeitig erkannt werden. Wie viele Kameras welche Daten über wie viele separate Eingänge am PC

gleichzeitig ihr Video schicken, sind wichtige Größen für die Menge der zu übertragenden Daten. Die Datenrate wird beeinflusst von

- Sensoranzahl (monokulare, binokulare, trinokulare etc. Systeme)
- Bildwiederholrate
- Auflösung
- Bittiefe pro Pixel [bit/px] bzw. [bpp] (Farbe, Quantisierung)
- Kompression (einige Industriekameras erlauben eine Kompression an der Kamera)

Bei den Hermes-Versuchen (Nr. 6) wurden von den Kameras Nr. 4 unkomprimiert von 2 der 3 verbauten Sensoren pro Kamera bei maximaler Bildwiederholrate und in voller Auflösung die Bilddaten monochrom mit 8 Bits Farbtiefe versandt. Damit ergibt sich eine Datenrate von

$$2 \cdot 16 \cdot 1280 \cdot 960 \cdot 8 \text{ bit} \approx 315 \text{ Mbit/s}.$$

Für die 170 Versuche mit einer durchschnittlichen Dauer von 98 s ergibt sich nur für die beiden eingesetzten Stereokameras ein Gesamtdatenvolumen von

$$2 \cdot 170 \cdot 98 \text{ s} \cdot 315 \text{ Mbit/s} \approx 1312 \text{ GB}.$$

Nach dem Übertragungsbus kommen die Bilddaten je nach Schnittstelle an einer entsprechenden PC-Karte an, die über den PC-internen Bus und den Hauptspeicher die Daten an die Steuereinheit der angeschlossenen Massenspeicher (Festplatte (ggf. RAID), SSD) leitet. Ist das Schreiben auf den Massenspeicher die Engstelle im Datentransfer und ist die jeweilige Versuchsdauer begrenzt klein, so kann das Wegschreiben auf den Massenspeicher und das Erfassen des Versuches entkoppelt und die Bilddaten im Hauptspeicher gepuffert werden. Die Auslastung des Prozessors liegt an der verwandten Schnittstelle und einer ggf. softwareseitigen Videokompression. Die Kompression kann sinnvoll sein, wenn die dem PC-internen Bus nachgeordneten Systemkomponenten eine zu geringe Durchsatzrate besitzen. Dabei muss gewährleistet sein, dass die Kompression verlustlos ist oder zumindest die Qualität der Bilddaten für die folgende Videoanalyse ausreichend bleibt.

Beispielsweise sind die maximalen Datenraten der für Versuchsreihe Nr. 6 beteiligten Komponenten:

- IEEE 1394b Übertragung bei Verlängerung auf 20 m mit zertifiziertem Kabel und aktivem Verstärker [548]: 800 Mbit/s
- IEEE 1394b PC-Karte: 2×800 Mbit/s
- PCI Express-Bus: 4000 MBit/s (2.0, Breite 1)
- Festplatte RAID 0 (Beschleunigung ohne Redundanz): 768 Mbit/s

Damit war gewährleistet, dass der Videostrom einer Stereokamera ohne Kompression mit 315 Mbit/s problemlos aufgezeichnet werden konnte. Auch wenn rechnerisch und bei Tests der Durchsatz der Komponenten eines PC für zwei Stereokameras gereicht hätte, wurde jede Aufzeichnung an einem separaten Rechner durchgeführt, weil eine Synchronisation auch über Rechnergrenzen hinweg möglich

war (siehe Abb. 4.13) und die Rechnerkosten im Vergleich zu den Gesamtkosten der Versuchsdurchführung gering waren. Zudem sind die Netto-Übertragungsraten je nach Anwendung geringer als die Brutto-Übertragungsraten [549]. Es muss darauf geachtet werden, dass keine belastenden Prozesse die Aufzeichnung stören (u. a. Virens Scanner).

Da gerade bei Laborversuchen die Gesamtkosten für einen Versuch inklusive der Kosten für die Probanden groß ist, ist die Sicherheit vor Datenverlust wichtig und kann durch direkte Vervielfältigung der Daten z. B. auf eine angeschlossene externe Festplatte erhöht werden. Dies wurde für die großen Versuchsreihen Nr. 3, 6 und 9 direkt vor Ort zum frühestmöglichen Zeitpunkt durchgeführt, um u. a. bereits Folgen durch Transportschäden zu verringern. Bei allen anderen Versuchsreihen geschah dies unmittelbar im Anschluss an die Versuchsdurchführung.

4.1.1.10 Montage

Bei Seitenaufnahmen (Versuchsreihe Nr. 1, 4, 5 und 7) wurden die Kameras auf Stativen aufgestellt, die auch für sonstige Aufnahmen von Übersichtskameras eingesetzt wurden. Die Montage der überkopf angebrachten Kameras und Beleuchtung erfolgte durch selbst verschweißte Schraubzwingen (Versuchsreihe Nr. 3, Abb. 4.14 a), zum Teil selbst gebaute Traversen (Versuchsreihe Nr. 2, 6 und 9, Abb. 4.14 c), an Fenstern befestigte Saugnäpfe (Versuchsreihe Nr. 8, Abb. 4.14 e und f) oder Kamerakräne (Versuchsreihe Nr. 6, Abb. 4.8 b). Auf den vorangegangenen Bildverweisen sind die Kameras grün umkreist, wenn sie aufgrund der geringen Größe der Industriekameras schwer zu erkennen sind. Im Stadion konnten die Laufstege im Stadionsdach nur für die Übersichtskameras genutzt werden, da sie zu weit entfernt waren. Die der Ermittlung der Laufwege dienenden Industriekameras wurden stattdessen an Kamerakränen befestigt, die auf eigens dafür gebaute Podeste aufgestellt wurden. In Abb. 4.8 sind am linken und oberen Bildrand zwei dieser Podeste zu erkennen.

Das Ziel war immer, die Kameras möglichst hoch zu positionieren, um unter Beachtung der Erkennung der Marker einen großen Bereich einsehen zu können oder durch Einsatz von Teleobjektiven die perspektivische Verzerrung und Verdeckung gering zu halten. Die Aufhängenhöhe variierte dabei zwischen 350 cm (Versuchsreihe Nr. 2), 531 cm – 539 cm (Versuchsreihe Nr. 3), 768 cm – 784 cm (Versuchsreihe Nr. 6 in den Messehallen), 245 cm – 1475 cm (Versuchsreihe Nr. 6 in der Arena), 2033 cm – 2044 cm (Versuchsreihe Nr. 8) und ca. 750 cm (Versuchsreihe Nr. 9).

4.1.2 Probanden und Aufbauten

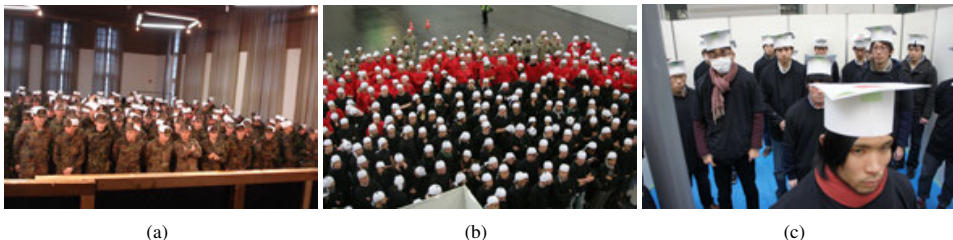


Abb. 4.15: Versuchsteilnehmer der Versuchsreihen (a) Nr. 3, (b) Nr. 6 und (c) Nr. 8

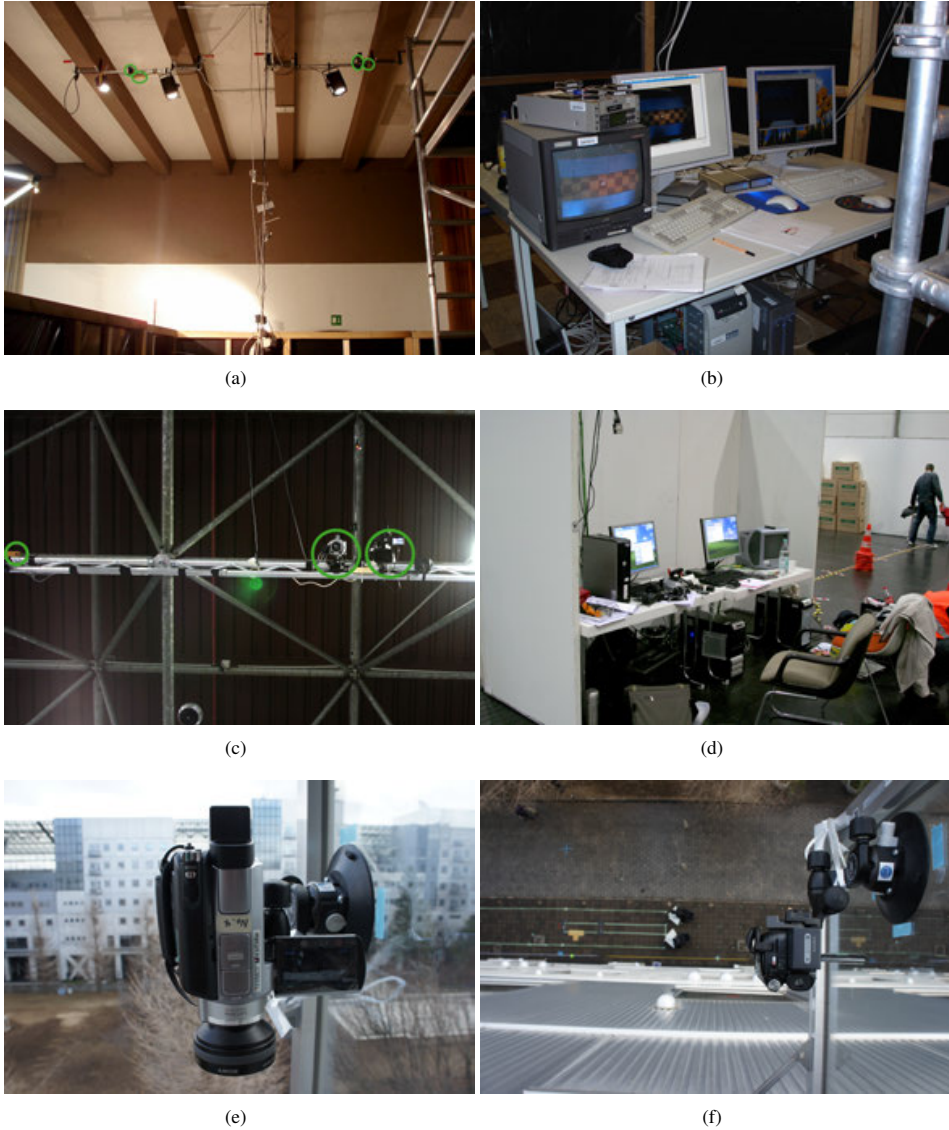


Abb. 4.14: Versuchsreihe Nr. 3: (a) Kameras (grün umkreist) und Aufhängung und (b) Arbeitsplatz mit Erfassungsrechnern; Versuchsreihe Nr. 6: (c) Kameras (d) Rechner zur Versuchsaufzeichnung und Überwachung; Versuchsreihe Nr. 8: (e, f) an Scheibe montierte Kamera in über 20 m Höhe

Für alle Versuche aus Tab. 4.1 waren die Probanden zum Großteil Studenten, da die Teilnehmer an Universitäten rekrutiert wurden (z.B. Altersstruktur der Hermes-Versuche (Nr. 6) war $25 \pm 5,7$ Jahre). Nur bei den Kasernen-Versuchen (Nr. 3) handelte es sich um Soldaten. Es wurden unterschiedlich detaillierte Fragebögen zur Datenerfassung ausgeteilt. Die umfassendste Befragung wurde bei den Ba-SiGo-Versuchen (Nr. 9) durchgeführt, da es dort das erste Mal möglich war, Individuen durch einen Code zu identifizieren und somit die Dynamik des Einzelnen direkt mit den persönlichen Eigenschaften in Verbindung zu bringen. Bei den umfangreichen Versuchsreihen Nr. 3 und 6 wurde u. a. die Größe erfragt. Ein Vergleich mit den automatisch ermittelten Werten ist in Abschnitt 3.6.2 zu finden.

Die Probanden mussten die jeweiligen Marker auf dem Kopf tragen. Für den Farbmarker (Nr. 4) (siehe Abb. 3.8 und 3.9) musste meist nur eine entsprechend farbige Kopfbedeckung aufgesetzt werden. Der Hermes-Marker (Nr. 2) war ebenfalls fertig auf Mützen (Bandanas) aufgebracht, die nur noch angezogen werden mussten (siehe Abb. 4.15 b). Der Kasernen-Marker (Nr. 1) wurde auf eine auf Kopfgröße zu faltende DIN-A4-Pappe gedruckt, die mit Büroklammern an der Mütze der Uniform der Soldaten angebracht wurde (siehe Abb. 4.15 a). Für den Japan-Marker (Nr. 3) wurde ein kompletter Hut in Form eines Doktorhutes erstellt, auf den die wegen der weiten Entfernung der Aufnahme von über 20 m besonders große Markerfläche (DIN A4) aufgebracht wurde.

Die Markerflächen auf Basis eines glatten bedruckten Papiers (Marker Nr. 1, 3 und 5) hatten den Vorteil, dass die Helligkeit der Fläche homogen war, dafür war der Tragekomfort bei reinen Kopfbedeckungen größer (Marker Nr. 2 und 4).

Neben der Kopfbedeckung wurden die Probanden zudem angewiesen, bestimmte Kleidung zu tragen, wenn ansonsten die Erkennung der Marker gestört wäre oder die Farbe der Oberhemden bei der manuellen visuellen Analyse die Personengruppen besser unterscheidbar machte (siehe u. a. Abb. 4.8).

Die künstlichen Aufbauten in den Laborversuchen wurden durch Leitkegel (Versuchsreihe Nr. 6 und 7, siehe Tab. 4.1), Stühle (Versuchsreihe Nr. 1, 3, 4 und 5), Seile (Versuchsreihe Nr. 1, 4 und 5), Tische (Versuchsreihe Nr. 2 und 3), Stell- oder Messewände (Versuchsreihe Nr. 5, 6, 8 und 9), selbstgebaute Wände (Versuchsreihe Nr. 3) oder Absperrgitter (Versuchsreihe Nr. 9) realisiert (siehe auch Abb. 4.1, 4.2 und 4.3). Dabei war darauf zu achten, dass neben der Bodenfarbe auch die Farbe der Aufbauten nicht die Erkennung der Marker stört. Hierzu wurden z. B. entsprechende Materialien zur Wanderstellung herangezogen oder die Tische mit passenden Tischdecken bedeckt (Versuchsreihe Nr. 3).

Für Stereoaufnahmen war es zudem wichtig, dass die Boden- und Wandbeschaffenheit keine spiegelnden Lichtreflexionen zuließen, da die Reflexe abhängig vom Betrachterstandpunkt wandern und damit die Disparitäts-Berechnung (siehe Abschnitt 3.4.2) falsche Werte liefern würde. Ebenfalls ist für Stereoaufnahmen von Vorteil, wenn die sichtbaren Objekte eine Struktur aufweisen, um eine möglichst durchgängige Disparität im ganzen Bild bestimmen zu können. Dass beide Bedingungen nicht voll erfüllt wurden, ist in Abb. 3.44 a und c gut zu erkennen. Der Lichtreflex auf dem Boden am linken Bildrand führte zu einem leichten Fehler in der Disparität an dieser Stelle, was für das auf der Disparität aufsetzende Verfahren nicht problematisch war. Störender war die strukturlose Wand, die zu sehr ungenauen Disparitäten insbesondere bei im Bild horizontal laufenden Wänden führte. Wie dies gelöst wurde, ist in Abschnitt 3.4.2 nachzulesen

4.2 Auswertung

Alle Versuche wurden videotechnisch festgehalten. Die Aufnahmen dienen zum einen der anschließenden manuellen visuellen Analyse und zum anderen der automatischen Extraktion der Laufwege.

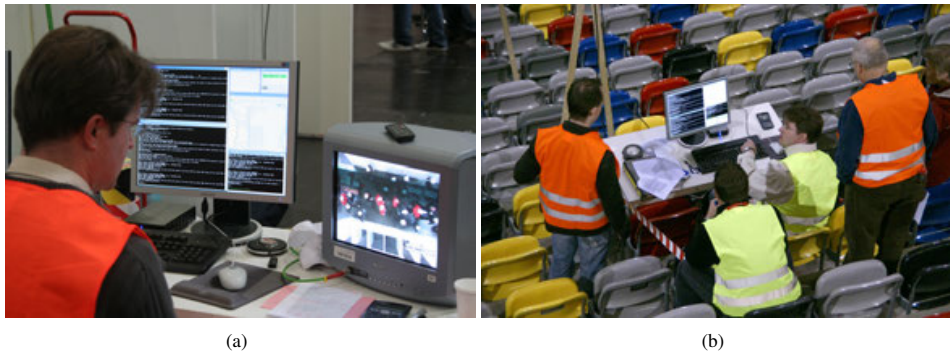


Abb. 4.16: Arbeitsplatz zur Steuerung und Überwachung der Versuchsaufzeichnung während der Hermes-Versuche (Nr. 6) (a) in den Messehallen und (b) in der Arena

Da die Aufnahmen sowieso zur späteren Überprüfung herangezogen werden, um oft auch erst wesentlich später ermittelte Effekte oder Gesetzmäßigkeiten zu validieren, und die Laufwege nicht während der Experimente benötigt wurden, war die Verfahrensentwicklung nicht auf eine Echtzeitfähigkeit hin ausgelegt. Dass es trotzdem Gründe für echtzeitfähige Verfahren gibt, ist Abschnitt 3.6.8 und Kapitel 5 zu entnehmen.

Bei den ersten Versuchsreihen Nr. 1 und 4 mit einer kleinen Anzahl an Personen und Durchläufen wurden die Aufnahmen händisch ausgewertet. Dabei wurden jedoch nur die Durchgangszeiten der Personen an wenigen Stellen (für gewöhnlich die Ein- und Austrittszeiten in bzw. aus dem Messbereich) festgehalten, um die Zeit der Erfassung im Rahmen zu halten. Hierfür wurden entweder real oder nachträglich in Videoschnittprogrammen Stäbe (siehe Abb. 4.1 b) bzw. Hilfslinien (siehe Abb. 4.1 a) aufgestellt bzw. eingezeichnet und personenbezogen die Durchgangszeiten in Tabellen erfasst. Dabei wurde immer vereinfacht davon ausgegangen, dass die Linien von allen Personen an der gleichen Position durchquert werden, also von der Seite die Personen im gleichen Abstand zur Kamera die Linie passieren und von oben alle Personen die gleiche Größe besitzen.

Für die halbautomatischen Verfahren, die bei den Versuchen Nr. 2 und 5 eingesetzt wurden, mussten die Personen von Hand markiert werden, wurden dann aber automatisch verfolgt. In der Versuchsreihe Nr. 2 geschah dies noch mittels der kommerziellen Software Adobe After Effects [550], für die Versuchsreihe Nr. 5 existierte bereits die selbst entwickelte Software *PeTrack*, die neben einer vollautomatischen Erfassung auch das reine Tracking von markierten Objekten erlaubt. *PeTrack* war damit bereits früh in der Lage halbautomatisch auch unmarkierte Personen zu verfolgen. Seit der Möglichkeit der halbautomatischen Auswertung waren detaillierte Trajektorien für die mikroskopische Analyse verfügbar.

Die weiteren Versuche Nr. 3, 6 und 8 sind vollautomatisch verarbeitet worden. Hierzu wurde die Software *PeTrack* entwickelt, die je nach Randbedingung (u. a. Personendichte, Abdeckung bzw. Entfernung zwischen Erfassungstechnik und Probanden, Geometrie (z. B. Treppen), Laborbedingung oder Feldstudie) seit dem Jahre 2006 fortentwickelt wird. Die Anpassung ist u. a. auch an den unterschiedlichen Markern in Abb. 3.8 und 3.9 zu erkennen. Die verschiedenen Verfahren werden ausführlich in Abschnitt 3 beschrieben. Eine Beschreibung der Verwendung auch der grafischen Oberfläche von *PeTrack* ist in Anhang D.3 zu finden. Für die automatische Auswertung einer ganzen Versuchsreihe wurde jedoch die Möglichkeit genutzt, die Software auch von der Kommandozeile aus steuern zu

können. Hierzu wurde für jede Kamera und ihrem Blickpunkt eine Kalibrierung (siehe Abschnitt 3.1) vorgenommen und in einer Projektdatei abgelegt. Mit dieser Projektdatei konnten nun mittels folgender Kommandozeile alle Versuche aus derselben Kameraeinstellung mittels Script zum Durchlaufen der einzelnen Versuche XXX ausgewertet werden:

```
petrack.exe -project projekt.pet  
            -sequence sequenzXXX  
            -autoTrack trajektorienXXX  
            -autoSave projektXXX.pet
```

Dabei erzeugt die Option `autoTrack` zum einen eine Datei, die alle Informationen der im Programm ermittelten Trajektorien insbesondere der Pixelkoordinaten und damit alle Daten für eine spätere Betrachtung der Ergebnisse des automatischen Verfahrens enthält (`trajektorienXXX.trc`) und zum anderen Dateien mit Trajektorien im realen Raum für die weitere Analyse, wobei verschiedene Ausgabeformate unterstützt werden.

Die Option `autoSave` erzeugt eine Projektdatei, die im Nachgang den Versuch XXX mit der Einzelbildsequenz oder Video `sequenzXXX` und den ermittelten Trajektorien `trajektorienXXX` direkt aufrufen lässt. Da für die anschließende Analyse keine Fehler in den Laufwegen enthalten sein sollen, sollte das Ergebnis manuell durchgeschaut werden. Je einfacher der Marker strukturiert war (z. B. einfacher Marker: Nr. 2; komplexer Marker: Nr. 1 in Abb. 3.8) und die Umgebung ähnliche Strukturen aufwies, desto höher war die Fehlerrate (siehe Abschnitt 3.6). Die Durchsicht nach potentiellen Fehlern wird von `PeTrack` dahingehend unterstützt, dass potentielle Fehlerstellen fortlaufend angesprungen werden können und die problematische Person hervorgehoben wird. Dabei sind mögliche Problemstellen die, bei denen z. B. zwei Trajektorien sehr nah beieinander liegen, die Geschwindigkeit innerhalb einer Trajektorie stark schwankt, eine Trajektorie innerhalb des Bildausschnittes beginnt oder endet und nicht durch den ganzen Versuchsaufbau hindurch läuft oder wenn die Länge einer ermittelten Trajektorie sehr kurz ist. Die Durchsicht wird auch dadurch erleichtert, dass alle Details einer bestimmten Trajektorie oder aller Trajektorien angezeigt werden können: die Positionen der Person in wählbaren Zeitintervallen, die Laufwege, Markerelemente, Personengrößen einschließlich Stellvertreterfarben und der individuellen Nummer (siehe Abb. D.5).

Wurde ein Versuch mit überlappenden Kameraperspektiven aufgenommen (Versuchsreihe Nr. 3, 6 und 8), so müssen im Anschluss die Trajektorien im realen Raum verschmolzen werden. Dies geschah mit einem ebenfalls selbstentwickelten separaten Programm `combine`, dessen Nutzung in Anhang D.4 beschrieben wird. Die Verfahrensweise ist in Abschnitt 3.3.5 nachzulesen.

Nach der halbautomatischen oder automatischen Extraktion der Trajektorien liegen die Daten für die weitere inhaltliche Analyse vor. Wie diese Analyse durchgeführt wurde und welche Ergebnisse damit erzielt wurden werden in den folgenden beiden Abschnitten beschrieben.

4.3 Messmethode

Zum Studium der Fußgängerdynamik werden beschreibende Größen benötigt. Dies sind vor allem die Geschwindigkeit v , die Dichte ρ und der davon ableitbare Fluss J . Die Dichte beschreibt dabei wieviele Personen pro Fläche Platz finden und der Fluss wieviele Personen pro Zeit eine Stelle passieren. Die Abhängigkeiten der Größen können im sogenannten Fundamentaldiagramm $J(\rho)$ oder $v(\rho)$ dargestellt werden (siehe beispielsweise Abb. 1.4). Die Korrelation ergibt sich u. a. daraus, dass die mögliche Geschwindigkeit einer Person mit Zunahme der Dichte sinkt, um eine Kollision zu vermei-

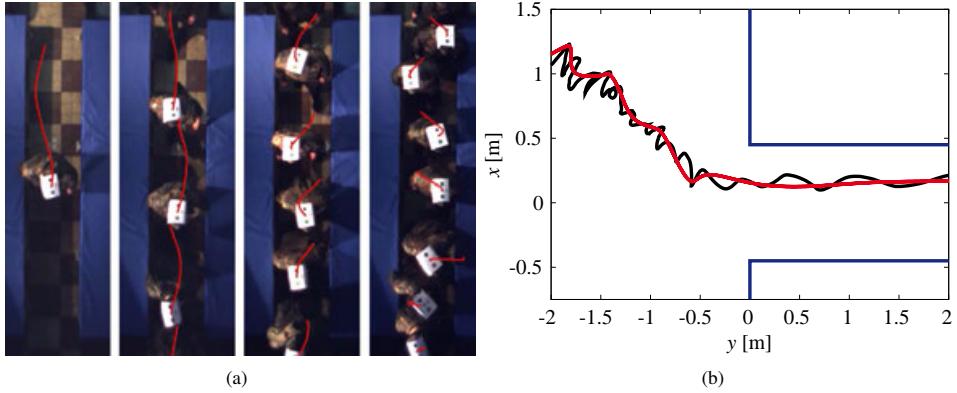


Abb. 4.17: Glättung von Trajektorien aufgrund geschwindigkeitsabhängiger Schwankung: (a) Die rote Linie zeigt den Weg der Personen in der letzten Sekunde bei eindimensionalem Gang durch einen Korridor; der Weg wird bei steigender Dichte von links nach rechts kürzer und die Bewegung entgegen der Hauptbewegungsrichtung größer; (b) Rohe (schwarz) und geglättete (rot) Trajektorie einer Person beim Durchqueren einer Engstelle mit Stauung; Bild (b) basiert auf Bildern aus [551]

den (siehe Video Nr. 23). Die Trajektorien liefern alle Informationen zur Bestimmung dieser Größen auch auf mikroskopischer Ebene und bieten dabei tieferen Einblick in die Bewegungen als einfache Durchgangszeiten.

Der Betrag der Geschwindigkeit $v_i(t)$ zum Zeitpunkt t einer Person i kann direkt aus den Trajektorien mit einem zeitlichen Abstand der zugrundeliegenden Einzelbilder von Δt gewonnen werden:

$$v_i(t) = \frac{\|p_i(t + \Delta t) - p_i(t - \Delta t)\|}{2\Delta t}. \quad (4.10)$$

Die Trajektorien zeigen bei den Überkopfaufnahmen den Weg des Kopfes. Für die Untersuchung der Dynamik in einer Menschenmenge ist jedoch die Hauptbewegungsrichtung bzw. Bewegung des Schwerpunktes von Interesse. Die obige Geschwindigkeitsdefinition würde die Schwankung einer Person insbesondere bei langsamer Bewegung (siehe Abb. 4.17 a und Video Nr. 23) und auch sonstige Kopfbewegungen ohne Bezug zur Gehrichtung (z. B. im Stehen) mit berücksichtigen. Dies würde dazu führen, dass die Geschwindigkeit und damit der Fluss immer überschätzt wären.

Daher werden die rohen Trajektorien p_i vor der weiteren Analyse geglättet, indem die Wendepunkte der Trajektorie durch einen Spline interpoliert werden (siehe Abb. 4.17 b). Die Wendepunkte in der Bewegung entsprechen den Übergängen bei der Belastung von einem auf das andere Bein. Die Schritt-erkennung durch Vorzeichenänderung der Krümmung erlaubt die Ermittlung der Schwerpunktbewegung. Mit der geglätteten Trajektorie \tilde{p}_i wird nun gemäß Gleichung 4.10 die geglättete Geschwindigkeit $\tilde{v}_i(t)$ bestimmt. Das Ergebnis ist insbesondere für Geschwindigkeiten oberhalb von 0,2 m/s zufriedenstellend [552]. Andere Methoden zur Glättung wären die Erhöhung von Δt in Gleichung 4.10, eine Frequenzanalyse oder die Berücksichtigung der Kopfausrichtung auf die Ausrichtung des Kopfes. In Abschnitt 3.6.2 wird ebenfalls auf die Schrittbewegung Bezug genommen und gezeigt, dass auch der Größenverlauf einer Person den Schritten folgt.

Zur Definition der Dichte gibt es verschiedene Methoden. Die hier eingesetzte Voronoi-Methode [552]

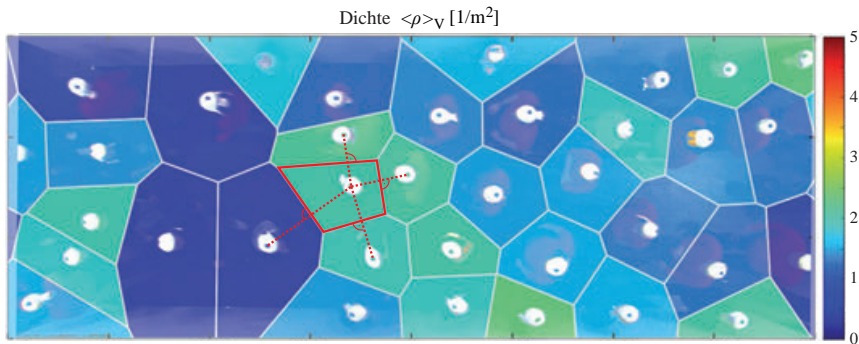


Abb. 4.18: Voronoi-Diagramm eines Korridor-Experiments; die Voronoi-Zellen sind entsprechend der Voronoi-Dichte eingefärbt; die rot umrandete Voronoi-Zelle besitzt vier Nachbarzellen, deren gemeinsame Kanten die jeweiligen Verbindungslinien der Personen mittig senkrecht teilt und so den der Person nächsten Bereich umschreibt

zeichnet sich dadurch aus, dass sie eine räumlich und zeitlich hohe Auflösung hat, eine geringe Fluktuation aufweist und für jede Position definiert ist. Sie basiert auf der eindeutigen Zerlegung der Fläche in Voronoi-Zellen $A_i(t)$ [553], die den Bereich beschreiben, der zur Zeit t der Person i am nächsten ist. Die Voronoi-Zelle wird zusätzlich durch die Beschränkung der Fußgängeranlage (z. B. Wände) und einen Kreis maximaler Ausdehnung begrenzt und kann somit als der persönliche Platzbedarf einer Person aufgefasst werden (siehe Abb. 4.18). Die Dichte- und Geschwindigkeitsaufteilung der Fläche ist damit

$$\rho(x, y) = \begin{cases} 1/A_i(t) & : \forall (x, y) \in A_i(t), \\ 0 & : \text{sonst}, \end{cases} \quad (4.11)$$

$$v(x, y) = \begin{cases} \tilde{v}_i(t) & : \forall (x, y) \in A_i(t), \\ 0 & : \text{sonst}. \end{cases} \quad (4.12)$$

Die Voronoi-Dichte und -Geschwindigkeit sind dann definiert als

$$\langle \rho \rangle_V = \frac{\iint \rho(x, y) dx dy}{\Delta x \cdot \Delta y}, \quad (4.13)$$

$$\langle v \rangle_V = \frac{\iint v(x, y) dx dy}{\Delta x \cdot \Delta y}. \quad (4.14)$$

Der spezifische Fluss ergibt sich daraus als

$$J_s = \langle \rho \rangle_V \cdot \langle v \rangle_V. \quad (4.15)$$

Der untersuchte Bereich $\Delta x \cdot \Delta y$ kann beliebig klein gewählt werden und ermöglicht damit einen detaillierten räumlichen Einblick. Die Fluktuation der Dichte ist bei der Voronoi-Definition geringer als bei der klassischen Methode, bei der die Anzahl der Personen in einem Bereich gezählt werden [552] (siehe Video Nr. 4). Das Video Nr. 55 zeigt die Voronoi-Zerlegung bei einem Korridor-Experiment und der Möglichkeit der variablen Raumaufteilung bis hin zum Profil des Versuchs, das im Folgenden erläutert wird.

Um ein Bauteil für ein gewisses Szenario zu bewerten ist oftmals nicht die Situation zu einem bestimmten Zeitpunkt von Interesse, sondern wie es sich für einen stationären Zustand auf Dauer verhält.

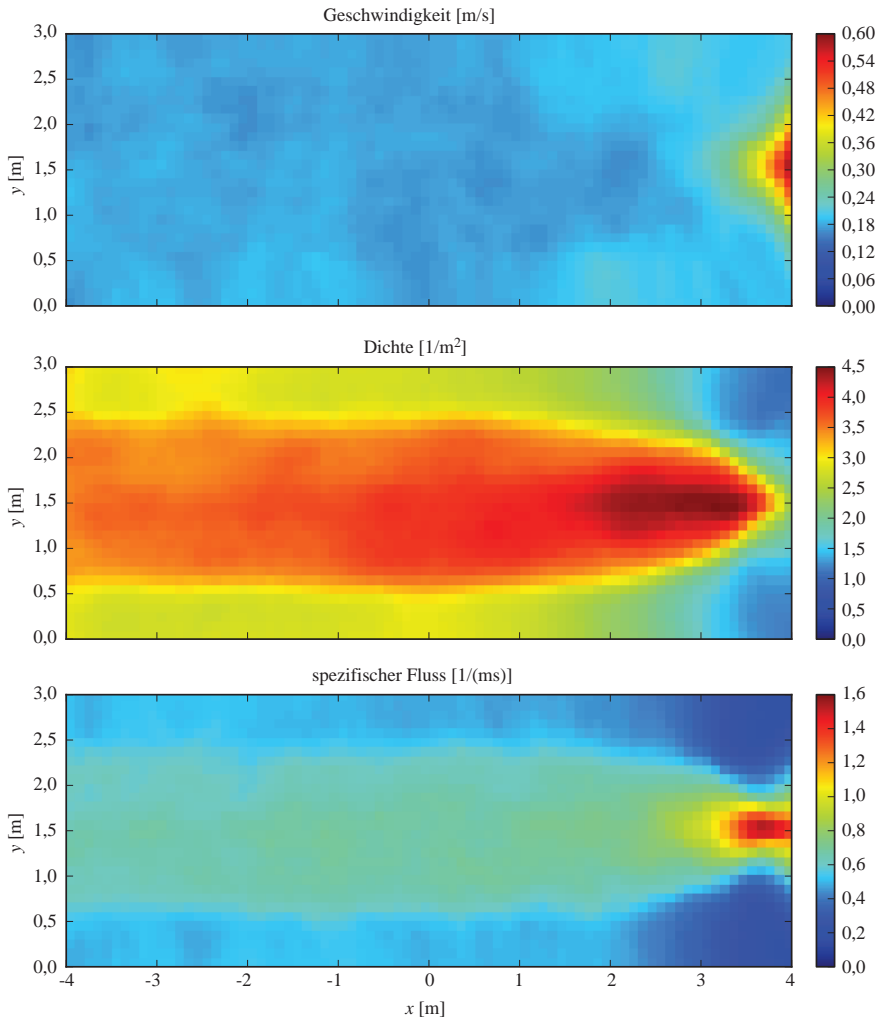


Abb. 4.19: Geschwindigkeits-, Dichte- und Fluss-Profil eines 3 m breiten Korridors mit einer Einschnürung von 80 cm; basiert auf Bild aus [554]

Der stationäre Zustand ist dabei der Zeitraum, in dem sich im Gesamtsystem die Geschwindigkeit und Dichte kaum verändern. Hierzu wird die Fläche des Bauteils fein parzelliert und gemäß Gleichung 4.13, 4.14 und 4.15 die interessierenden Größen für jeden erfassten Zeitschritt während des stationären Zustandes bestimmt und über die Zeit gemittelt. Für eine äquidistante Aufteilung eines Versuches in einem 3 m breiten Korridor mit anschließender Einschnürung auf 80 cm in Kacheln der Größe $\Delta x \times \Delta y = 10 \text{ cm} \times 10 \text{ cm}$ ergeben sich Geschwindigkeits-, Dichte- und Fluss-Profile gemäß Abb. 4.19 [554]. Die Profile geben neue Einblicke in die räumliche Charakteristik der den Personenverkehr beeinflussenden Werte. Die Bestimmung solcher Profile wie auch die zugrundeliegende Voronoi-Messung ist nur mithilfe von exakten Trajektorien möglich.

4.4 Anwendung

Nr.	Veröffentlichungen zur Erfassung ¹	Szenarien bzw. unters. Parameter	Veröffentlichungen ² (Vergleich mit Simulation)	Videos ³
1	[535]	Korridor	[536, 555–558], [17, 20, 559–564]	[26]
2	[18]	Engstellenbreite	[141, 142, 565–568], [55, 505, 561, 562, 564, 569, 570]	[8, 14, 29, 38]
3	[56, 57, 571, 572]	Korridor	[472, 573], [560, 561, 574–578]	[23, 24]
		Engstelle	[19, 472, 552, 579–582], [55, 505, 551, 561, 569, 575, 576, 578, 583–587]	[1, 2, 4, 16, 17, 18, 36, 50]
4	[536]	Kultur	[557]	[26]
5	[537]	Motivation		[41]
6	[31, 32, 56, 57, 572, 588]	Tribünen, Treppen	[526, 527, 589, 590], [532, 583, 591, 592]	[39, 43, 44, 45, 46, 47, 48, 42]
		Korridore	[21, 554, 593–596], [505, 569, 597]	[7, 54, 55, 25]
		Ecken, Kreuzungen	[31, 594, 598, 599]	[25, 30, 31, 32, 33, 34, 35]
7	[600]	Fahrradverkehr	[601, 602]	[19, 20, 21]
8		Personenschlange	[508, 603]	[27]
		Raumein-, austritt	[508, 604]	[28]
9		kritische Dichten		[5, 6, 13, 12, 10, 9, 11]

Tab. 4.3: Liste der auf den Experimenten beruhenden Veröffentlichungen, bei denen unsere Gruppe beteiligt war; die Nummerierung bezieht sich auf die Nummer der Versuchsreihe in Tab. 4.1;

¹Veröffentlichungen, die die Erfassung der Laufwege am ausführlichsten beschreiben;

²bei kursiven Referenzen werden die Daten zur Modellierung herangezogen (Entwicklung, Kalibrierung, Validierung);

³verweist auf die zu diesem Thema in Anhang C aufgeführten Videos

In diesem Abschnitt sollen exemplarisch Auswertungen und Anwendungen vorgestellt werden, die Laufwege als Eingabe verwenden, die mit den in dieser Arbeit vorgestellten Verfahren ermittelt wurden. Dabei wurde die Analyse bzw. Bewertung der Ergebnisse meist nicht von mir, sondern von Kollegen durchgeführt. Trotzdem sind sie in diese Arbeit aufgenommen worden, um aufzuzeigen, welche Möglichkeiten der Analyse exakte Laufwege bieten.

Tab. 4.3 listet die Veröffentlichungen bis Mitte 2014 auf, die Laufwege der in Tab. 4.1 aufgeführten Versuchsreihen nutzen. Sie beschränkt sich auf Veröffentlichungen unter Beteiligung des Forschungszentrum Jülich. Dabei wird in der Tabelle unterschieden, ob eine Veröffentlichung die Erfassung der Laufwege ausführlich beschreibt, eine reine Analyse beinhaltet oder aber die Daten zur Modellbildung, -kalibrierung oder -validierung herangezogen wurden. Die Arbeiten beschäftigen sich u. a. mit der Dichte-Fluss-Relation [472], dem Einfluss der Motivation [537], der Kultur [536, 557], der Engstellenbreite [19, 565–567, 579, 580, 582], der Engstellenlänge [19, 579], der Treppensteigung

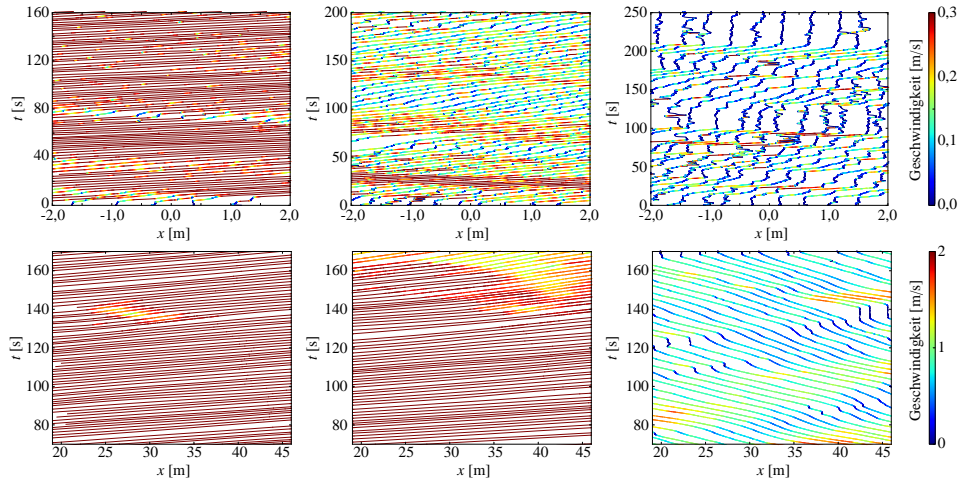


Abb. 4.20: Raum-Zeit-Diagramme von eindimensionalen Bewegungen in Hauptbewegungsrichtung x ; (oben) Personen aus Versuchsreihe Nr. 3 bei Dichten von 1,5/m, 2,15/m und 2,69/m; das Auftreten und der Verlauf von Stauwellen sowie die Adaption der Bewegung ist ablesbar; (unten) Fahrradfahrer aus Versuchsreihe Nr. 7 bei Dichten von 0,23/m und für die letzten beiden Diagramme von 0,38/m; die leichte Stauwelle im letzten Diagramm im Gegensatz zum vorherigen trat nur bei der Bewegung im Uhrzeigersinn auf; weitere Informationen siehe [600]

[532, 590], der Bewegungsrichtung [21, 554, 595] und speziellen Geometrien wie einer T-Kreuzung [31, 594, 598, 599]. Darüberhinaus gibt es eine Vielzahl von weiteren Veröffentlichungen, die unsere Daten aufgrund der freien Bereitstellung [533] genutzt haben, z. B. [141, 605, 606].

Mit Durchgangszeiten von Personen an ausgezeichneten Messstellen können zwar lokale Flüsse erfasst werden (z. B. Abb. 4.22), die Geschwindigkeit oder Dichte aber nur global zwischen diesen Messstellen bestimmt werden, wenn eine Zuordnung der Personen möglich ist. Mit Trajektorien ist die Information über Geschwindigkeit, Dichte und Fluss zu jeder Zeit an jeder Stelle verfügbar. Zudem können weitere Bewegungsstrukturen analysiert werden wie z. B. das geschwindigkeitsabhängige Schwankungsverhalten (siehe Video Nr. 23), der Verlauf einer Stauwelle (siehe Abb. 4.20) oder die Ausnutzung der Fläche durch die Personen (siehe Abb. 4.19 und 4.32). Darüber hinaus kann mit speziellen Markern die Kopfausrichtung (Kasernen- (Nr. 1) oder Japan-Marker (Nr. 3)) erfasst oder die Laufwege mit erhobenen Daten der Probanden (Codemarker (Nr. 5)) verknüpft werden, um z. B. den Einfluss des Geschlechts oder Laufmuster von sozialen Gruppen untersuchen zu können.

4.4.1 Engstelle

Ein häufig vorkommendes und für die Sicherheit wichtiges Element in Fußgängeranlagen sind Engstellen wie z. B. Türen. Sie bewirken durch die Verringerung des Flusses eine Erhöhung der Personendichte und können dadurch ggf. kritische Situationen erzeugen.

Engstellen wurden auf vielfältige Weise in den Versuchsreihen Nr. 2, 3, 6, 8 und 9 untersucht. Der Einfluss der Breite, der Länge, der initialen Dichte sowie von Wellenbrechern vor der Engstelle und von Eingliederungsverfahren wurde studiert. Abb. 4.22 zeigt den Einfluss der Engstellenbreite und -länge auf den maximal in den Experimenten ermittelten Fluss durch die Engstelle anhand unserer Versuchs-

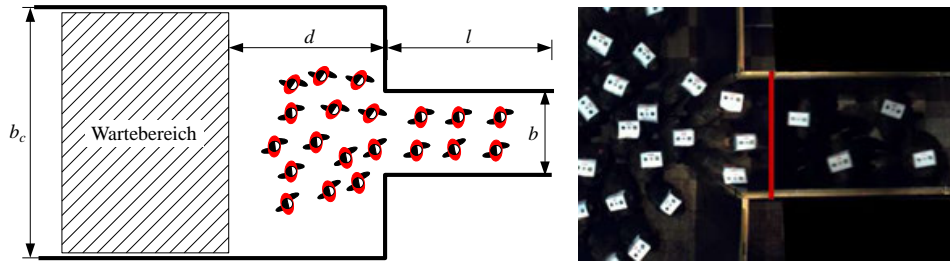


Abb. 4.21: (links) Aufbauskizze der Engstellenversuche aus Versuchsreihe Nr. 3 und (rechts) aus zwei Kameraperspektiven zusammengesetzte Aufnahme eines Experimentes

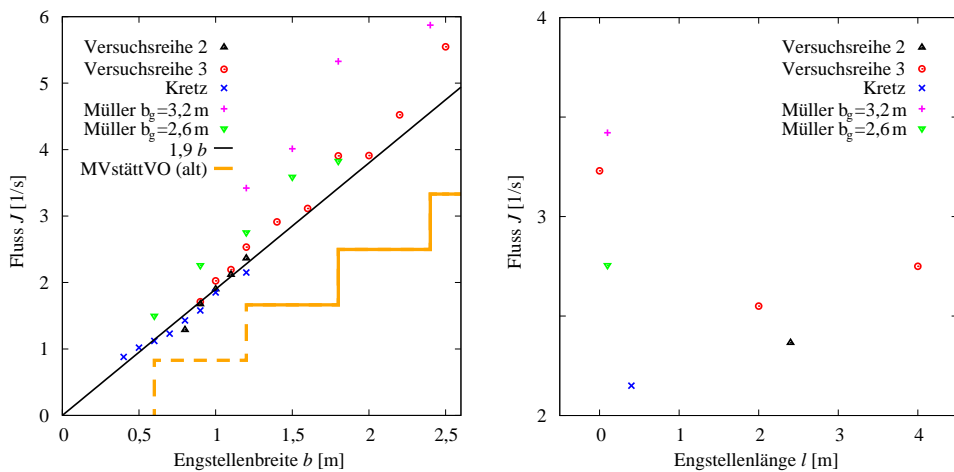


Abb. 4.22: Abhängigkeit des maximal in den Experimenten ermittelten Flusses J innerhalb einer Engstelle von ihrer Breite b und Länge l ; die Daten stammen von den Versuchsreihen Nr. 2 und 3 sowie aus [116, 607]; (links) die alte Musterverordnung der Versammlungsstättenverordnungen der deutschen Bundesländer (MVStättVO (alt)) gab eine stufenweise Erhöhung des maximalen Flusses im Verhältnis zur Breite vor, die experimentellen Daten zeigen einen linearen Zusammenhang von ca. $J(b) = 1,9b$ /(ms); (rechts) für kurze Engstellenlängen ist der maximale Fluss signifikant höher; weitere Informationen siehe [57, 576]

reihen Nr. 2 und 3 (siehe Abb. 4.21) sowie Daten aus Kretz [116] und Müller [607]. Diese Erhebung wäre bereits mit einer Messstelle in der Engstelle realisierbar, jedoch liefern erst die Trajektorien eine Begründung für den beobachteten Zusammenhang.

Die Abhängigkeit des maximalen Flusses J von der Breite b ist ungefähr linear und kann mit $J(b) = 1,9b$ /(ms) approximiert werden (siehe Abb. 4.22 links). Die Daten von Müller liegen signifikant höher, möglicherweise aufgrund einer sehr hohen Startdichte von $6/\text{m}^2$ vor der Engstelle. Die von der Fachkommission Bauaufsicht der Bauministerkonferenz erlassene alte Musterverordnung der Versammlungsstättenverordnungen der deutschen Bundesländer (kurz: Musterversammlungsstättenverordnung, MVStättVO) sah jedoch in der Vergangenheit eine stufenweise Erhöhung der erforderlichen Ausgangsbreiten vor und gab eine Flusststeigerung von $0,83/\text{s}$ alle 60 cm an [5, 6] (siehe Abb. 4.22 links). Durch Umsetzung unserer Eingaben zur Änderung der Musterversammlungsstättenverordnung entsprechend den oben genannten Ergebnissen eines linearen Zusammenhangs ist die Planung von

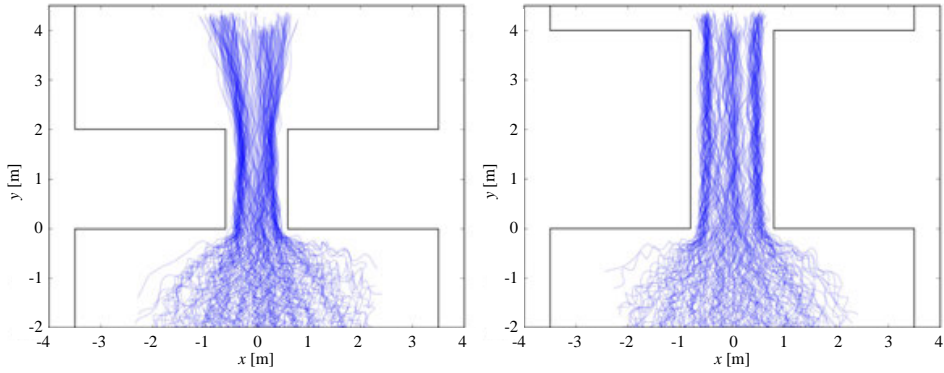


Abb. 4.23: Laufwege durch zwei Engstellen unterschiedlicher Breite (1,2 m, 1,6 m) und Länge (2 m, 4 m); eine Linienbildung ist ersichtlich, aber auch die Nutzung des Raumes zwischen den Linien; weitere Informationen siehe [57, 579]

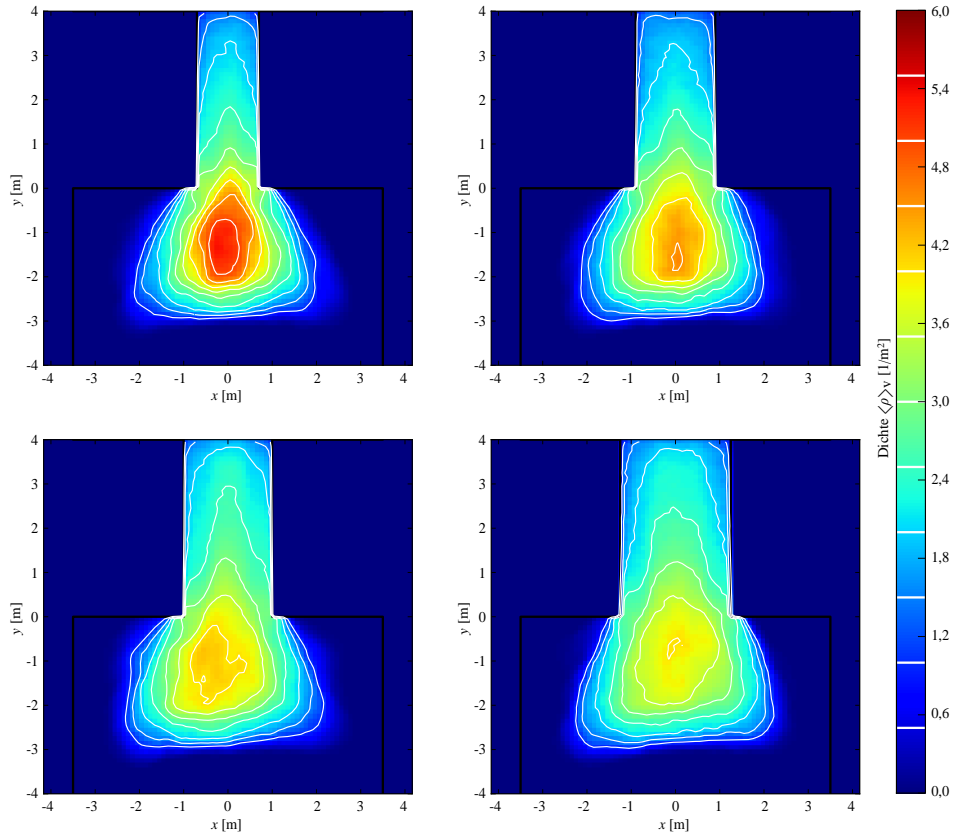


Abb. 4.24: Dichte-Profil von Engstellen der Breite 1,4 m, 1,8 m, 2,0 m sowie 2,5 m und einer Länge von 4 m; weiße Linien sind Isolinen gleicher Dichte im Abstand von 0,5/m²; die höchste Dichte liegt mittig ca. 1 m vor der Engstelle und wandert bei steigender Breite an die Engstelle heran; weitere Informationen siehe [19, 596]

Versammlungsstätten nun realitätsnäher und kann u. a. aufgrund des entfallenden Rasters der Türbreiten flexibler durchgeführt werden. Die die Verbesserungen enthaltende aktuelle Fassung der Muster-versammlungsstättenverordnung vom Februar 2014 muss noch in die für Nordrhein-Westfalen gültige Sonderbauverordnung (SBauVO) aufgenommen werden. Die Idee hinter der stufenweisen Erhöhung ist die auch in Abb. 4.23 zu erkennende Linienbildung innerhalb von Engstellen und der Annahme, dass sich der Fluss erst dann erhöhen kann, wenn sich eine weitere Linie ausgebildet bzw. die Breite sich schulterbreit erhöht hat. Bei den überlagerten Trajektorien ist jedoch gut zu erkennen, dass der Raum zwischen den dominierenden Linien bei Erhöhung der Breite, hier von 1,2 m auf 1,6 m (laut Musterversammlungsstättenverordnung wäre der gleiche maximale Fluss von 1,7/s anzusetzen gewesen), immer stärker ausgenutzt wird (siehe Video Nr. 1 und 8). Bei sehr breiten Engstellen verschwimmen die Linien immer weiter (siehe Video Nr. 2) und nur die beiden äußersten Linien sind noch erkennbar.

Ein weiterer Aspekt, der erst mit Hilfe der Laufwege untersucht werden konnte, ist die Lokalisierung hoher Dichten innerhalb einer Fußgängeranlage, also den Bereichen, die für Menschen bei hohem Personenaufkommen besonders gefährlich sind. Die Dichteverteilung bei variierender Engstellenbreite kann Abb. 4.24 entnommen werden. Sie zeigt, dass für eine Engstelle der Ort höchster Dichte mittig ca. 1 m vor der Engstelle liegt. Der Abstand zur Engstelle verringert sich dabei leicht, je breiter die Engstelle wird.

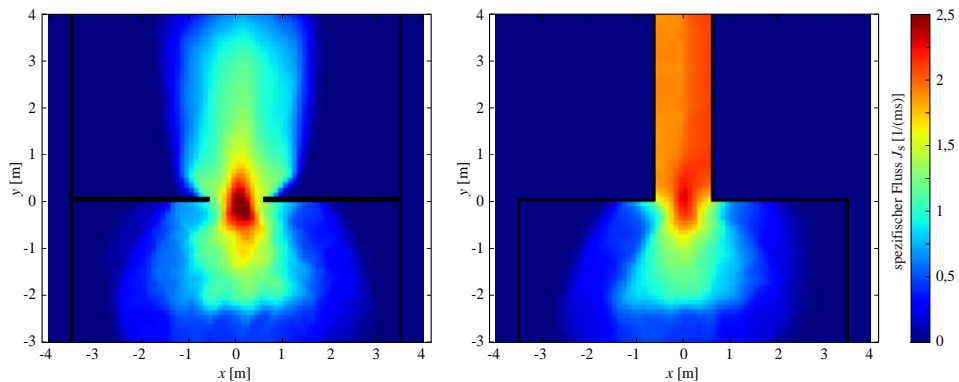


Abb. 4.25: Fluss-Profil einer Engstelle der Breite 120 cm und einer Länge von 6 cm und 400 cm; der Fluss ist in der kurzen Engstelle aufgrund der größeren Geschwindigkeit (siehe Abb. 4.26) höher; weitere Informationen siehe [19]

Auch die Engstellenlänge hat einen Einfluss auf den maximal beobachteten Fluss bei den Experimenten. Abb. 4.22 zeigt einen größeren Fluss für kurze Engstellen. Das Fluss-Profil für die Längen 6 cm und 400 cm ist Abb. 4.25 zu entnehmen. Innerhalb der kurzen Engstelle ist der Fluss größer, da die Geschwindigkeit signifikant höher liegt und obwohl die angenommene Dichte geringer ist (siehe Abb. 4.26). Bei der Dichte ist zu bedenken, dass die an die Engstelle anschließenden Wände des Korridors bei der Voronoi-Methode starken Einfluss nehmen, da die Voronoi-Zellen durch diese beschnitten werden. Eine Erklärung für den Unterschied des Flusses könnte sein, dass Menschen eine temporäre Annäherung eher zulassen als eine höhere Dichte über einen längeren Bereich entlang eines angeschlossenen Korridors und damit in einer kurzen Engstelle diese schnell passieren, um danach die Nähe auflösen zu können (siehe Video Nr. 36).

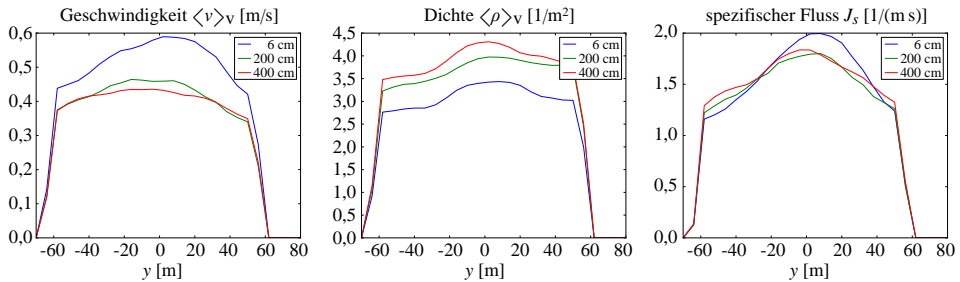


Abb. 4.26: Geschwindigkeit, Dichte und spezifischer Fluss im Querschnitt am Beginn einer Engstelle der Breite 120 cm und den Längen 6 cm, 200 cm und 400 cm; der signifikante Unterschied für die kürzeste Engstelle und die ähnlichen Verläufe der beiden längeren Engstellen ist erkennbar; weitere Informationen siehe [19]

4.4.2 T-Kreuzung

Im Rahmen des Projektes Hermes [527, 608] wurden Versuche (Versuchsreihe Nr. 6) durchgeführt, in denen u. a. zusammenfließende Ströme untersucht wurden. Hierzu wurden sieben Experimente an einer T-Kreuzung mit variierenden Öffnungsbreiten des Zustroms realisiert (siehe Abb. 4.27 und 4.28).

Die Laufwege nehmen, wie in Abb. 4.29 zu sehen, je nach Größe des Zuflusses unterschiedliche Bereiche der Fläche ein. Daraufhin wurde das Fundamentaldiagramm für die einzelnen Zweige der Kreuzung separat bestimmt und es zeigt sich, dass der vereinigte Personenstrom einen zu den Zuströmen unterschiedlichen Dichte-Geschwindigkeits-Zusammenhang besitzt (siehe Abb. 4.30).

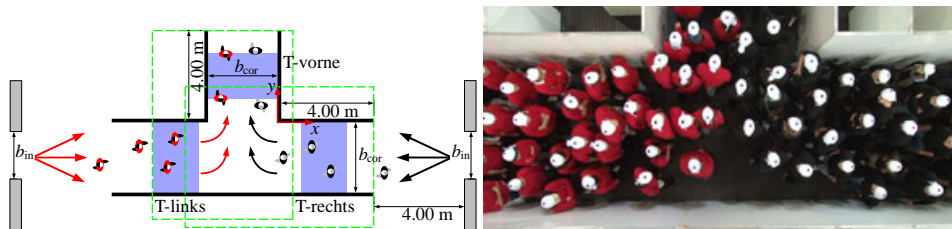


Abb. 4.27: Aufbauskeizze der T-Kreuzungs-Versuche aus Versuchsreihe Nr. 6 und senkrechte Aufnahme eines Experimentes



Abb. 4.28: Schrägsicht auf den experimentellen Aufbau durch Messestellwände eines T-Kreuzungs-Versuches

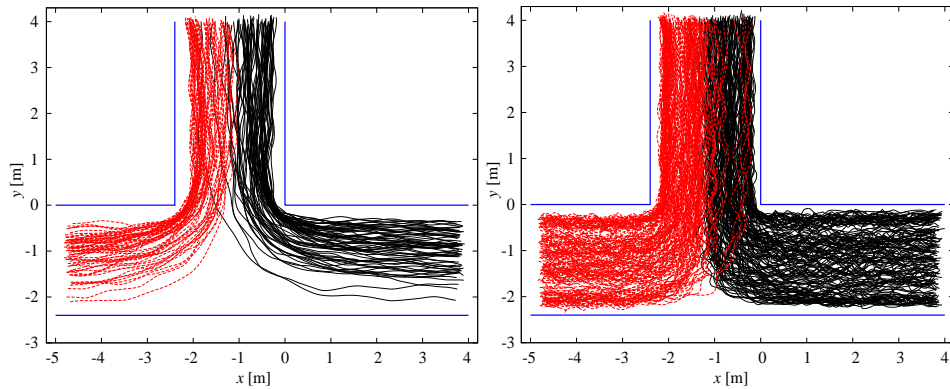


Abb. 4.29: Laufwege zusammenfließender Personenströme an einer T-Kreuzung mit einer Korridorbreite von $b_{cor} = 2,4$ m; die Breite des Zuflusses wurde auf $b_{in} = 0,5$ m bzw. $2,4$ m begrenzt; die Flussrichtung und die Bezeichnungen sind Abb. 4.27 zu entnehmen; die Nutzung des Raumes variiert mit den einströmenden Personenmengen; weitere Informationen siehe [57]

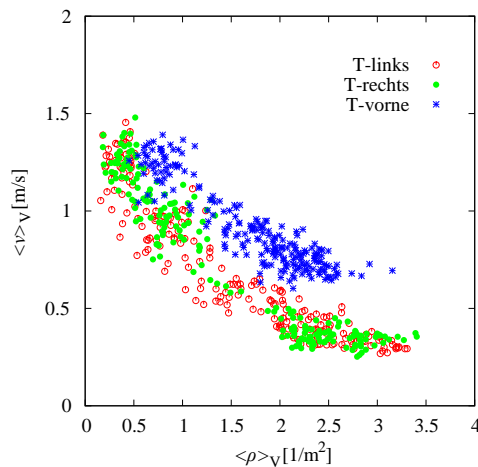


Abb. 4.30: Fundamentaldiagramm als Dichte-Geschwindigkeits-Diagramm der in Abb. 4.27 markierten Messbereiche T-links, T-rechts und T-vorne einer T-Kreuzung; der Unterschied zwischen den Zuströmen (T-links und T-rechts) und dem vereinigten Personenstrom (T-vorne) ist klar zu erkennen; weitere Informationen siehe [31, 599]

Der Unterschied könnte daher rühren, dass Personen, die mit einer Stauung rechnen oder diese vor sich sehen, die Geschwindigkeit bereits im Voraus absenken. Bei Annahme eines ungehinderten Flusses im Korridor, der aus dem Versuchsaufbau herausführt, wird eine höhere Geschwindigkeit bei gleicher Dichte bevorzugt.

Diese Beobachtung zeigt, dass ein gutes Modell der Fußgängerdynamik nicht nur versuchen darf, ein universelles Fundamentaldiagramm abzubilden. Abb. 4.31 zeigt den Vergleich einer Simulation mit experimentellen Ergebnissen anhand eines Versuchs in einer T-Kreuzung mit einer Zuflussbreite von $b_{in} = 1$ m. Die Simulation wurde mit der Software FDS+Evac [609] durchgeführt und stellt ein weiteres Beispiel einer Simulation dar, die wesentliche Größen wie Dichte, Geschwindigkeit und Fluss sowie die Nutzung des Raumes nicht korrekt wiedergibt. Im folgenden Abschnitt wird beschrieben, wie unser Modell für die Anwendung in einem Evakuierungsassistenten kalibriert wurde.

Das Dichte-, Geschwindigkeits- und Fluss-Profil mit dem maximalen Personenaufkommen in der Kreuzung (Zuflussbreite entspricht der Korridorbreite von $b_{in} = b_{cor} = 2,4$ m) sind Abb. 4.32 zu entnehmen. Der Bereich größter Dichte liegt unmittelbar vor dem ausgehenden Korridor, wo sich die Personenströme vereinigen. Nach der Vereinigung wird direkt der maximale Fluss erzielt und die Geschwindigkeit sukzessive erhöht, was eine Verringerung der Dichte zur Folge hat.

Eine Überkopfaufnahme eines Versuches mit maximaler Zuflussbreite ist in Video Nr. 32 zu sehen.

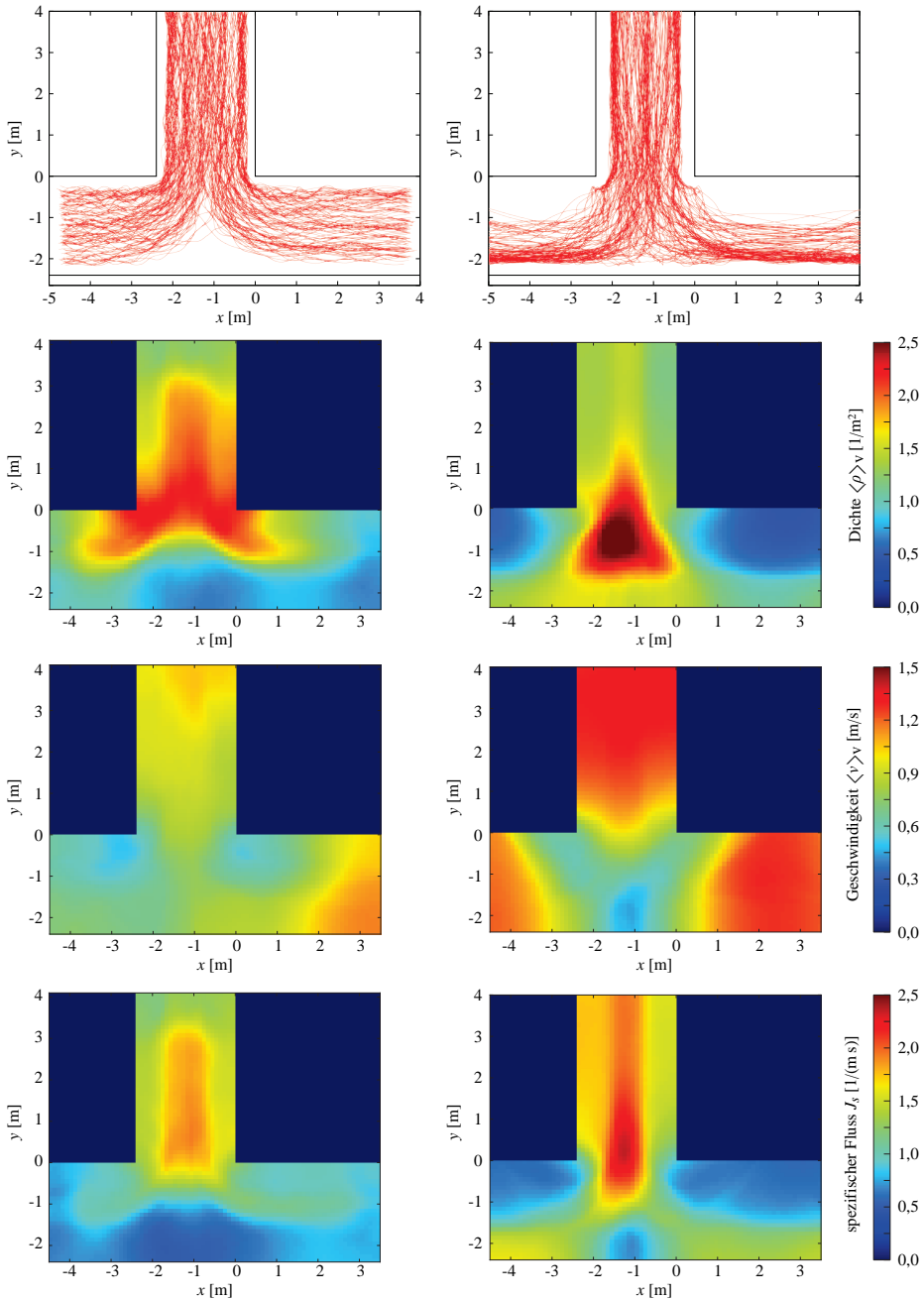


Abb. 4.31: Laufwege sowie Dichte-, Geschwindigkeits- und Fluss-Profil einer $b_{cor} = 2,4$ m breiten T-Kreuzung mit einer Zuflussbreite von $b_{in} = 1$ m; (links) experimentelle Daten der Hermes-Versuche; (rechts) Simulationsergebnisse ermittelt mit der Software FDS+Evac [609]; Unterschiede in der Nutzung des Raumes sowie allen Profilen sind signifikant; basiert auf Bildern aus [532]

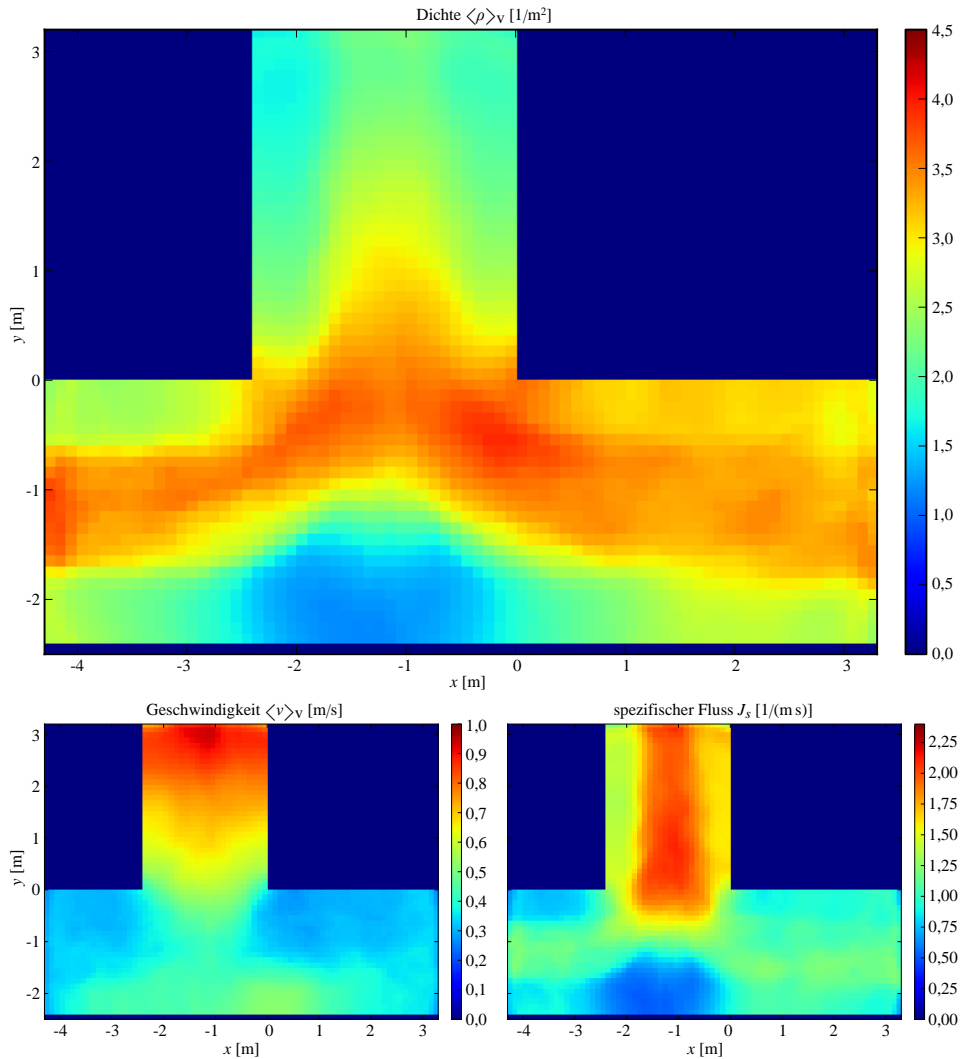


Abb. 4.32: Dichte-, Geschwindigkeits- und Fluss-Profil einer $b_{cor} = 2,4$ m breiten T-Kreuzung mit einer Zuflussbreite von $b_{in} = 2,4$ m; weitere Informationen siehe [31, 594, 598, 599]

Video Nr. 30 zeigt die Ermittlung der Laufwege, die in Video Nr. 35 mittels der Software `travisto` [610] visualisiert werden. In Video Nr. 34 wird der Prozess der markerlosen Extraktion für einen T-Kreuzungs-Versuch maximaler Dichte demonstriert, indem die getragenen Marker ignoriert werden. Die in jedem Zeitschritt ermittelten Voronoi-Zellen und entsprechenden Dichten und Geschwindigkeiten sind in Video Nr. 31 visualisiert. Das Resultat einer Simulation der Personenbewegung durch eine T-Kreuzung wird im Video Nr. 33 dargestellt.

4.4.3 Evakuierungsassistent

Im Rahmen des Projektes Hermes [527, 608] ist ein Evakuierungsassistent für die ESPRIT Arena in Düsseldorf entwickelt worden, der den Entscheidungsträgern wie Betreiber, Polizei, Feuerwehr und Sicherheitsdienst im Krisenfall neben Informationen über den Zustand des Bauwerks und die aktuelle Verteilung der anwesenden Personen auch eine Prognose über den Evakuierungsverlauf bzw. eine Vorhersage über die Entwicklung der Besucherströme in den kommenden 15 Minuten lieferte. Auf Grundlage dieser Informationen war es den Einsatzleitern vor Ort möglich, die Gefahrenlage frühzeitig einzuschätzen und somit Rettungskräfte und Sicherheitspersonal optimal einzusetzen.

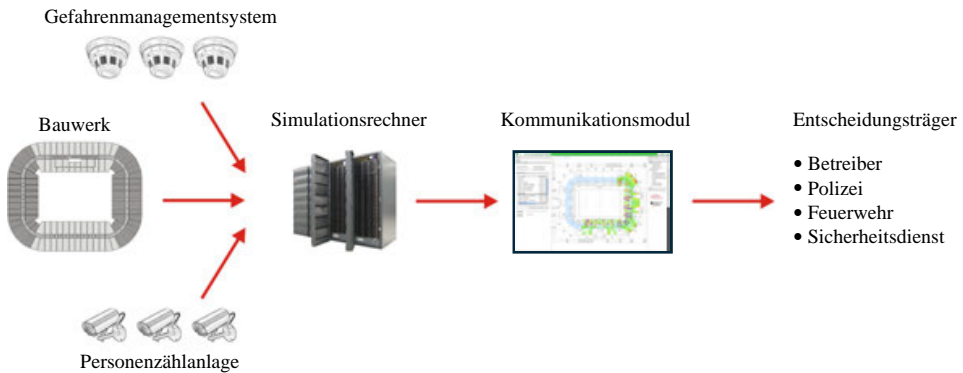


Abb. 4.33: Schema des Evakuierungsassistenten; die zuvor oder aktuell erfassten Daten werden der Simulation zur Verfügung gestellt, die über ein Kommunikationsmodul (siehe Abb. 4.35 links) den Entscheidungsträgern eine Vorhersage präsentiert; weitere Informationen siehe [527, 591, 608]

Die Funktionsweise des Assistenten geht aus der Skizze in Abb. 4.33 hervor. Daten über die Gebäudestruktur, von dem Gefahrenmanagementsystem und von einem extra für den Assistenten zur automatischen Personenzählung installierten Netzwerk von Kameras wurden kontinuierlich an einen eigens dafür vorgesehenen Parallelrechner geliefert, der auf Anfrage der Einsatzleiter eine viertelstündige Vorhersage berechnete. Die Vorhersageergebnisse wurden über ein Kommunikationsmodul (siehe Abb. 4.35) den Entscheidungsträgern präsentiert, die zudem interaktiv Gebäudeparameter wie den Zustand einer Tür (offen oder geschlossen) verändern konnten, um den Einfluss solch einer Einflussnahme besser einschätzen zu können.

Die in Abschnitt 4.4.2 vorgestellten Versuche gehörten zu 170 Versuchen der Versuchsreihe Nr. 6, die zur Erforschung der Personenbewegung während eines Evakuierungsfalles durchgeführt wurden. Dabei wurden generelle Geometrien wie Kreuzungen und Ecken, aber auch für die ESPRIT Arena spezielle bauliche Gegebenheiten wie die Tribüne, Mundlöcher und Außentreppen untersucht, die im Falle einer Evakuierung genutzt würden. Video Nr. 45 zeigt, wie auch für nicht ebene Versuche die Extraktion mit Hilfe von Stereokameras durchgeführt wurde.

Das zur Simulation der Evakuierung neben einem Zellularautomaten-Modell (*cellular automata model*, CA) der Universität zu Köln [564, 616] eingesetzte raumkontinuierliche generelle Zentrifugalkraft-Modell (*generalized centrifugal force model*, GCFM) [55, 569, 584] beruht auf einem kraftbasierten Ansatz und bezieht anziehende Kräfte zum Ziel und abstoßende von benachbarten Personen und Wänden ein. Das Modell beschreibt den Körper und Platzbedarf der Fußgänger in der Bewegungsebene als Ellipsen, deren Größe von der Geschwindigkeit abhängt. Für die Kalibrierung des Modells konnten mit Hilfe der Laufwege zum einen der geschwindigkeitsabhängige Platzbedarf aufgrund der

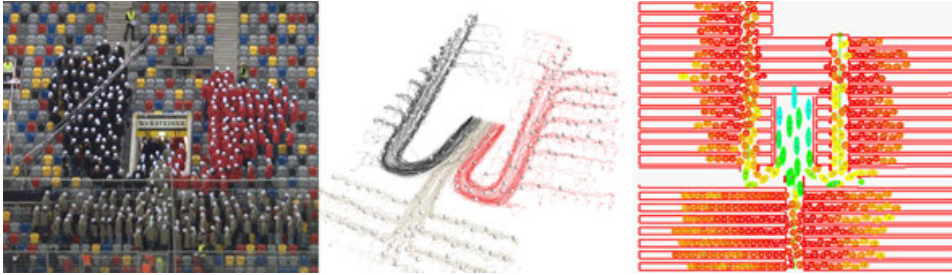


Abb. 4.34: Gegenüberstellung eines Versuchs beim Verlassen der Tribüne durch ein Mundloch, der ermittelten dreidimensionalen Laufwege und einer Simulation derselben Geometrie; der gesamte Versuch und der Vergleich mit einer Simulation ist Video Nr. 43 zu entnehmen; rechtes Bild basiert auf Arbeiten aus [611–615]

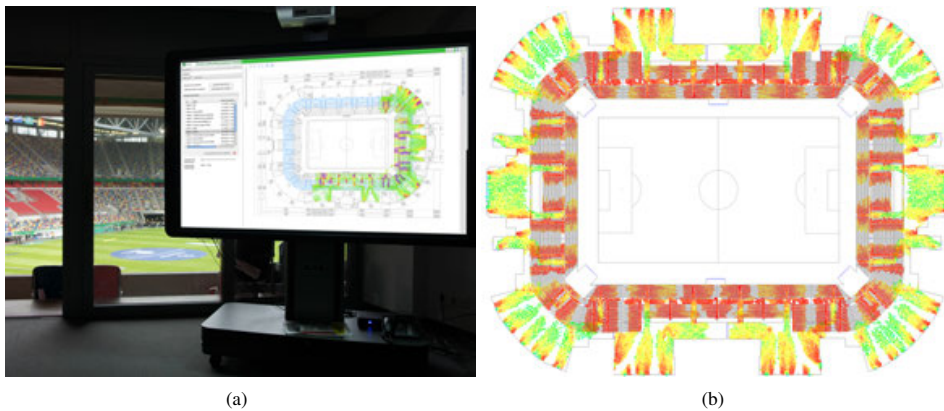


Abb. 4.35: (links) der Leitstand bzw. das Kommunikationsmodul des Evakuierungsassistenten in Form einer interaktiven Bedieneroberfläche auf einem großformatigen berührungsempfindlichen Bildschirm vor einem Fenster mit Blick auf den Zuschauerraum und das Spielfeld der Arena; (rechts) Simulationsergebnis der Vorhersage einer Evakuierung der gesamten Arena nach zehn Sekunden der Ankündigung der Räumung; die Farbe gibt die Geschwindigkeit wieder (rot $\hat{=}$ langsam bis grün $\hat{=}$ schnell), so dass rote Bereiche aufgrund der bei langsamer Bewegung ansteigenden Dichte kritischere Bereiche anzeigen; rechtes Bild basiert auf Arbeiten aus [611–615]

Schrittlänge und der Schwankung bestimmt und für die unterschiedlichen Gebäudegeometrien entlang des Fluchtweges die entsprechenden realen Dichte-Geschwindigkeits-Verhältnisse herangezogen werden. Die Profile der Personenbewegung werden für die Kalibrierung der Modelle herangezogen, um ihre Generalität zu testen. Eine für alle Fälle korrekte Vorhersage ist jedoch nicht möglich [592]. Die Gegenüberstellung eines Experiments und einer Simulation des gleichen Szenarios in Abb. 4.34 ist für den gesamten Versuch dem Video Nr. 43 zu entnehmen. Eine Visualisierung der Simulationsergebnisse zur Evakuierung der Promenade der ESPRIT Arena ist in Video Nr. 52 zu sehen.

Zusammenfassung und Ausblick

5.1 Zusammenfassung

Die in dieser Arbeit vorgestellten Techniken und entwickelten Verfahren haben es ermöglicht, präzise Laufwege von Personen in Menschenmengen hoher Dichte zu erfassen. Hierzu wurden Kameras oder Stereokameras eingesetzt, da sie auch für eine große Personenzahl genaue Ergebnisse liefern können. Sie besitzen die höchste Auflösung aller Sensoren und damit die meisten extrahierbaren Informationen zur Detektion einer Person. Es existieren eine Vielzahl von entwickelten Verfahren im Bereich der Bildanalyse, da Kameras und Stereokameras auch aufgrund ihres geringen Preises weit verbreitet sind. Für die Analyse der Personenbewegung erlauben Kameras überdies die manuelle Sichtung des Videomaterials, um Ergebnisse der automatischen Analyse verstehen, einordnen und bewerten zu können.

Alle Aufnahmen wurden überkopf mit senkrechter Ausrichtung zum Boden durchgeführt, um auch bei hohen Dichten, die hinsichtlich der Sicherheit besonders relevant sind, die geringste Verdeckung zu haben. Der dadurch eingeschränkte Sichtbereich wurde durch überlappende Perspektiven mehrerer Kameras erweitert, deren Laufwege durch zeitliche Synchronisation und örtliche Kalibrierung automatisch zusammengeführt werden konnten.

Zur Extraktion der präzisen Laufwege aus den Videoaufnahmen wurden alle Kameras zuvor kalibriert sowie der zur Elimination der perspektivischen Verzerrung benötigte Abstand zum erfassten Kopf jeder Person bestimmt. Die Bestimmung der einzelnen Kopfabstände wurde mit Hilfe von speziellen Markern oder dem Einsatz von Stereokameras insbesondere für Versuche an Treppen realisiert. Am Ende besaßen die ermittelten Laufwege je nach Winkel zur optischen Achse der Kamera und Entfernung zur Kamera einen maximalen Fehler in der Bewegungsebene von 0 cm direkt unterhalb der Kamera bis zu 5 cm in den Ecken des Sichtfeldes. Die Erkennung wurde in den Laborexperimenten mit Hilfe von Markern durchgeführt, die je nach Einsatzgebiet unterschiedlich gestaltet waren. Der Einsatz von Markern erlaubte erst eine robuste, immer denselben Punkt des Kopfes verfolgende Trajektorie zu erheben.

Der Kasernen-Marker konnte aufgrund der geringen Kameraentfernung zu den Köpfen der Versuchsteilnehmer komplex strukturiert sein, so dass es zu keiner falsch positiven (0%) und einer geringen Anzahl an falsch negativen ($< 1\%$) Erkennung einer Person kam und die Kodierung der Personen-

größe über den Marker zuließ. Die Erkennung des Markers erfolgte durch Suche nach entsprechend dem Aufbau des Markers angeordneten und gerichtete Isolinien gleicher Helligkeit approximierenden Ellipsen. Der ähnlich zu detektierende Hermes-Marker war einfacher aufgebaut, da aufgrund der Entfernung von bis zu 13 m an Treppen die Pixelabdeckung der Stereokameras pro Kopf geringer war, was zu falsch positiven (bis 4%) und wenigen falsch negativen ($< 1\%$) Detektionen führte. Durch Kombination der Markerelemente mit einem markerlosen Verfahren konnten für die Hermes-Versuche die falsch positiven Erkennungen stark reduziert werden ($\sim 0\%$). Beim Japan-Marker wurde eine größere, etwas vereinfachte Variante des Kasernen-Markers verwendet, da die Kopfausrichtung mit erfasst werden sollte. Hierbei kam es wie beim Farbmarker zu keiner einzigen Fehldetektion. Bei den Versuchsreihen mit diesen beiden fehlerfreien Markern wurde auf eine Erfassung der unterschiedlichen Körpergrößen verzichtet und stattdessen der entstehende Fehler durch eine hohe Kameraaufhängung von über 20 m reduziert. Der Codemarker wurde eingeführt, um individuelle Laufwege zu erfassen und benötigte dafür eine hohe Pixelabdeckung und entsprechend für die große Versuchsfläche eine Vielzahl an Kameras, deren Sichtfelder sich gegenseitig überlappen.

Zur Extraktion der Laufwege wurden die detektierten Personen durch Tracking über das gesamte Video verfolgt, wobei zur Qualitätssteigerung automatische Methoden entwickelt wurden, die u. a. eine Verbesserung der ermittelten Laufwege bei mehrmaligem Durchlaufen ermöglichen, fehlerhaftes Tracken kurzfristig durch Extrapolation überbrücken und fehlende Bilder in Aufnahmen der Industriekameras einfügen können.

Die Laufwege der fast 600 in den letzten neun Jahren durchgeführten Laborversuche bieten eine in ihrem Umfang und ihrer Qualität weltweit einmalige Datenbasis für die Analyse der Bewegung von Personen in Menschenmassen und führten bis heute zu mindestens 74 Publikationen allein der Arbeitsgruppe am Forschungszentrum Jülich, u. a. im Bereich der validierten Modellbildung für die Fußgängerdynamik. Die untersuchten Szenarien erstrecken sich über Korridore unterschiedlicher Länge, Breite und Bewegungsrichtung, Ecken, Engstellen, Kreuzungen mit diversen Zulaufrichtungen und Winkel, bis hin zu verschiedenen Treppentypen und -steigungen. Zudem erfolgte in Experimenten eine Variation des Kulturkreises und der Motivation der Versuchsteilnehmer. Die meisten Versuchsszenarien wurden über ein breites Intervall an Personendichten untersucht.

Mit der Vielzahl an Laborexperimenten konnte eine große Menge wichtiger Parameter studiert werden, welche die Dynamik einer Menschenmenge beeinflussen. Die Menge an möglichen Einflussgrößen ist jedoch so umfangreich, dass diese mit Laborexperimenten nicht erschöpfend untersucht werden kann. Es fehlen z. B. noch Versuche zum Einfluss der Streuung der Körpergrößen, der Lichtverhältnisse, der Kleidung oder des Gepäcks. Einige Faktoren sind nur unter natürlichen Bedingungen variabel oder lassen sich nur sehr schwer künstlich einstellen, so z. B. die Psyche der Versuchsteilnehmer oder der Einfluss von Rauschmitteln. Um die Erhebung von weiteren Daten zu vereinfachen und günstiger durchführen zu können, sollen in Zukunft ebenfalls Feldstudien durchgeführt werden, auch wenn die Randbedingungen gar nicht oder schlechter einzustellen und zu erfassen sind, wie z. B. die Zusammensetzung der Teilnehmer.

Für diesen Zweck wurde ein auch bei hohen Personendichten robustes Verfahren entwickelt, das die im Vergleich zu bisherigen Verfahren geringste Fehlerrate besitzt. Die Falsch-Ermittlungs-Rate liegt für das entwickelte Verfahren bei reinen Personenströmen auch bei hohen Personendichten von $4/m^2$ bei 0% und steigt z. B. am Eingangstor einer realen Großveranstaltung auf 2% aufgrund des Vorkommens von Objekten, die der Form des Kopfes oder Oberkörpers ähneln, wie z. B. mitgeführte Luftballons, da nur die Körperform und keine weiteren Merkmale für die Erkennung herangezogen werden. Die Falsch-Negativ-Rate lag um 1%. Da die Ermittlung der Körperform Störungen besitzt

und nicht mehr genau der Marker auf dem Kopf verfolgt werden kann, verringert sich aufgrund fehlender Marker die Glattheit der Laufwege.

Die Ermittlung der Körperform geschieht direkt über die perspektivische Disparitätsmatrix bzw. das Tiefenfeld der Stereokamera, deren optische Achse wie zuvor bei den monokularen Kameras senkrecht zum Boden ausgerichtet sein muss. Der die Personen enthaltende Vordergrund wird durch eine Hintergrundsubtraktion ermittelt und besteht aus den Pixeln, die der Kamera näher als das Tiefenfeld der leeren Szene sind. Das Tiefenfeld des Vordergrundes wird in Schritten von 5 cm bis 10 cm abgetastet und Isolinien gleicher Entfernung zur Kamera durch der Körperform ähnelnde Ellipsen approximiert, um einen einfacheren Zugriff u. a. auf Form und Orientierung der Isolinien zu haben. Ellipsen sich einschließender Isolinien werden zu sogenannten Ellipsenpyramiden zusammengefasst. Merkmale wie u. a. die Anzahl der in einer Pyramide enthaltenden Ellipsen, die Achsenlängen oder der Flächeninhalte markanter Ellipsen innerhalb der Pyramide werden untersucht und zur Identifizierung einer Person herangezogen.

Alle Entwicklungen an den Verfahren zur Ermittlung von Laufwegen sind in die freie Software *PeTrack* [456] eingeflossen, die bereits von vielen anderen Forschergruppen genutzt wird. Die ermittelten Laufwege werden nach eigener Analyse ebenfalls der Allgemeinheit für deren Untersuchungen frei zur Verfügung gestellt [533].

5.2 Ausblick

Die vorgestellten und implementierten Verfahren zur Laufwegebestimmung haben sehr gute Ergebnisse erzielt (siehe Abschnitt 3.6), trotzdem sind eine Vielzahl von Erweiterungen oder Verbesserungen denkbar, die die Extraktion der Trajektorien ggf. robuster, schneller oder genereller und die abgeleiteten Größen noch verlässlicher machen könnten.

Bereits während der Vorstellung der selbst entwickelten Verfahren wurden einige Ideen zur Weiterentwicklung angesprochen, aber auch darüber hinaus sind in den letzten Jahren der Entwicklung viele Ideen aufgekommen, die bislang noch nicht weiter verfolgt werden konnten. Einige sollen im Folgenden aufgeführt werden. Sie können die bereits existierenden Verfahren möglicherweise verbessern oder erweitern.

Der Farbmarker besaß ohne eine Fehldetektion mit die geringste Fehlerrate, zeigte aber bei der Positionstreuung des als Ort einer Person angenommenen Punktes auf dem Kopf systematische Fehler zum Rand und erzeugte unruhigere Laufwege als Marker mit Strukturelementen (siehe Abschnitt 3.6.4). Hingegen führte eine sehr einfache Struktur als Marker zu falsch positiven Detektionen ähnlicher Objekte (siehe Abschnitt 3.6.3). Sehr einfache Strukturen waren jedoch aufgrund der Entfernung zur Kamera und leichter Verdeckungen wegen hoher Dichten nötig (siehe Abschnitt 4.1.1.1). In einem aktuell begonnenen Forschungsprojekt sollen Personen mit sehr unterschiedlichen Körpergrößen bei hoher Dichte verfolgt werden. Hierzu könnten die positiven Eigenschaften des Farbmarkers und des einfach strukturierten Markers zu einem neuen Marker verbunden werden, dessen Strukturelement markante Farben besitzen könnte, die darüber hinaus zur Kodierung der Personengröße genutzt werden könnten.

Der dokumentierte systematische Fehler bei der Detektion des Farbmarkers und des markerlosen Verfahrens in Abhängigkeit vom Blickwinkel zur optischen Achse könnte herausgerechnet werden (siehe Abschnitt 3.6.4.2, 3.6.4.3 und 3.6.6.2). Hierzu sind weitere Analysen des Fehlers und Angaben von Nutzer nötig. Der Fehler beim Farbmarker aufgrund des Blickwinkels ist u. a. abhängig von der Art

der Kopfbedeckung (z. B. tief sitzende Mütze oder Schirmmütze mit großer Sonnenblende), beim markerlosen Verfahren wird der Fehler mit der gewählten Methodik zur Bestimmung der Disparität und der Schrittweite, mit der das Tiefenfeld abgetastet wird, variieren.

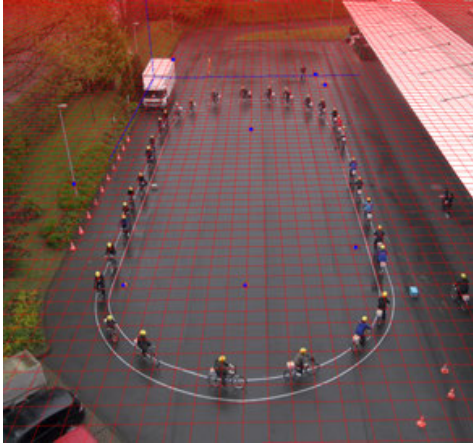


Abb. 5.1: Schrägsicht eines Fahrradexperimentes aus Versuchsreihe Nr. 7; das rote Gitter mit $1\text{ m} \times 1\text{ m}$ großen Gitterzellen ist in durchschnittlicher Kopfhöhe eingezeichnet; die blauen Punkte sind Kalibrierungspunkte zur Bestimmung der Lage des blau eingezeichneten Weltkoordinatensystems

Bei Einsatz einer Stereokamera kann die reale Position eines Pixels bei Kenntnis der internen Kameraparameter direkt aus der Disparität abgeleitet werden (siehe Gleichung 3.37), so dass keine exakt senkrechte Sicht nötig ist, was bereits bei der ersten Feldstudie genutzt wurde (siehe Abschnitt 3.6.7.2). Bei monokularen Kameras wird bisher von einer senkrechten Ausrichtung der optischen Achse zur Bewegungsebene ausgegangen. Dies ist zwar für die Minimierung von Verdeckungen sinnvoll, erschwert aber die flexible Nutzung der Verfahren. So könnte die Arbeit der exakten Ausrichtung der Kamera entfallen oder auch bei geringeren Personendichten mit leicht schrägen Kameraansichten gearbeitet werden, wenn die extrinsischen Kameraparameter der äußeren Orientierung, also die Abbildung vom Welt- in das Kamerakoordinatensystem, flexibler zu bestimmen und festzulegen wären. Zudem fehlen bei Feldstudien häufig Möglichkeiten, die Kameratechnik überkopf anzubringen. Erste Ergebnisse sind

in Abb. 5.1 zu sehen, bei dem unter Verwendung von [458, 617] ein Gitter mit $1\text{ m} \times 1\text{ m}$ großen Gitterzellen in durchschnittlicher Kopfhöhe in ein Bild eines Experimentes aus Versuchsreihe Nr. 7 gelegt wurde, bei dem die geringe Personendichte eine Schrägsicht zuließe. Die blauen Punkte stellen Kalibrierungspunkte dar. Das blau eingezeichnete resultierende Weltkoordinatensystem hat seinen Ursprung im oberen linken Bereich des Bildes.

Durch die senkrechte Ausrichtung der Kameras überkopf kam es bisher nur zu Teilverdeckungen des zu detektierenden Kopfes, so dass das Tracking durch Suche nach ähnlichen Pixelblöcken in Folgebildern erfolgen konnte. Bei Erweiterung des Sichtfeldes, dem Anstieg der Differenz der Körpergrößen oder einer in Zukunft erlaubten schrägen Ausrichtung der Kamera ist mit temporärer kompletter Verdeckung von Köpfen zu rechnen, so dass andere Verfahren zum Tracken herangezogen werden müssen, die Bewegungen schätzen und eine Wiedererkennung aufgrund von Form, Farbe oder Muster erlauben.

Neben den Detektionsmethoden mit Markern besitzt auch die markerlose Erkennung eine sehr geringe Fehlerrate (siehe Abschnitt 3.6.6.1). Durch das fehlerbehaftete Disparitätsfeld sind jedoch die Höhenlinien und damit die Lage der die Position festlegenden approximierenden Ellipsen starken Schwankungen ausgesetzt (siehe Abschnitt 3.6.6.3). Eine Verbesserung könnte entweder die Entwicklung neuer Methoden zur Bestimmung der Disparitäten herbeiführen, oder aber die in Abhängigkeit des Blickwinkels erkennbare Systematik der Lage der approximierenden pyramidenförmigen Ellipsenstapel (siehe Abschnitt 3.5.5) dafür genutzt werden, stabilere Positionen des Kopfes und damit auch ruhigere Laufwege zu erhalten. Hierzu könnten die Mittelpunkte der Ellipsen eines Ellipsenstapels durch eine dem Blickwinkel und der Lage in vorangegangenen Bildern entsprechenden Gerade

approximiert werden, die am Durchstoßpunkt der größten Erhöhung im Tiefenfeld einer Person die Position einer Person bestimmt.

Die Durchführung von Feldstudien ist insbesondere dort interessant, wo es häufig zu hohen Personendichten kommt. Dies ist z. B. bei Konzerten der Fall, die aber oft abends oder zur besseren Wirkung der Lichteffekte abgedunkelt durchgeführt werden (siehe links in Abb. 1.1) und damit den Einsatz von Stereokameras erschweren oder gar unmöglich machen. Das markerlose Verfahren nutzt die dreidimensionale Form des Oberkörpers einer Person zu seiner Erkennung (siehe Abschnitt 3.5). Die benötigte 3D-Information ist ebenfalls mit aktiven RGB-D-Sensoren wie der Kinect extrahierbar, die bei künstlichem Licht oder auch im Dunkeln einsetzbar ist (siehe Abschnitt 2.3.7). Aus diesem Grund soll in Zukunft das resultierende Tiefenfeld einer Kinect dahingehend untersucht werden, ob es wie das Tiefenfeld der Stereokamera für das vorgestellte markerlose Verfahren genutzt werden kann. Damit wäre zudem eine sehr kostengünstige Lösung zur Personendetektion gefunden, da eine Kinect nur ein Zehntel der bisher eingesetzten Stereokameras kostet. Des Weiteren erlaubt diese Art von Sensor eine Untersuchung des Einflusses der Sichtverhältnisse, die u. a. bei Evakuierungsprozessen nach einem Brand oder einem Stromausfall eingeschränkt sein können, auf die Bewegung [618, 619].

Heutige auch mikroskopische Modelle der Fußgängerdynamik berücksichtigen nicht das komplette individuelle Verhalten einer Person, da über dieses in der Menschenmenge wenig bekannt ist. Die ersten Versuche zur Erhebung individueller Verhaltensmuster bzw. -regeln wurden mit den BaSiGo-Versuchen (Nr. 9) unternommen, indem über individuelle Codes (siehe Abb. 3.9 5) und dazugehörige Fragebögen eine Aussage darüber gemacht werden kann, wie z. B. zusammengehörende Personen einer Gruppen sich je nach Größe unterschiedlich bewegen oder das Geschlecht Einfluss auf das Verhalten in einer Menschenmenge hat. Den Einfluss der Personengröße auf die Dynamik kann bereits mit den bisher erhobenen Daten untersucht werden. Genauere Modelle könnten aber auch die Blickrichtung und damit den Sichtbereich oder den Bewegungsablauf über die Schrittfolge berücksichtigen. Zur Erfassung der Blickrichtung in Menschenmengen müssten zusätzliche Sensoren wie Eyetracker hinzugezogen werden. Bereits vorliegende Informationen sind die Kopfausrichtung, die über die Marker Nr. 1 und 3 ausgelesen werden können (siehe Abb. 3.8 1 und 3). Nach einer Untersuchung der Korrelation der Blick- und Kopfausrichtung wäre auch über Kopfmarker eine Aussage z. B. über die Erfassung von Hinweisschildern oder die Fokussierung auf einen Ausgang realisierbar. Die exakte Modellierung der Bewegung von Personen bedürfte neben des Einflusses mentaler Faktoren der Berücksichtigung des kompletten Bewegungsablaufs des Bewegungsapparates. Für Bewegungen in Menschenmengen sind diese Daten in Gänze nur mittels zusätzlicher z. B. inertialer Sensoren zu erheben (siehe Abschnitt 2.3.10). Die wichtigste Bewegung neben der Hauptbewegungsrichtung ist die Schrittfolge und den damit einhergehenden Armbewegungen. Diese kann bereits mit den detailliert erhobenen Laufwegen bestimmt werden. Mögliche Detektionsmethoden der Schritte werden im folgenden Absatz angesprochen.

Vor der Bestimmung der für die Fußgängerdynamik wichtigen Größe der Geschwindigkeit in Hauptbewegungsrichtung einer Person wird im Moment zur Glättung eines Laufweges der Wendepunkt in der Trajektorie und damit in der Schrittfolge gesucht und diese Punkte mittels eines Splines interpoliert (siehe Abschnitt 4.3). Dies erzeugt bei wenig gestörten Bewegungen oberhalb von 0,2 m/s gute Ergebnisse. In Stausituationen sind andere Verfahren nötig. Eine Mittelung über die Zeit ist eine einfache Methode, die aber echte Richtungsänderungen ebenfalls egalisiert (wähle großes Δt in Gleichung 4.10). In langsamen Bewegungen sind die Wendepunkte beim Verlagern des Körpers von einem auf das andere Bein schwer in der Trajektorie zu lokalisieren, somit könnte die mittels Stereokamera detektierte Auf- und Abbewegung (siehe Abschnitt 3.6.2) eine stabilere Erkennung der Schrittfolge erlauben oder diese zumindest unterstützen. Eine andere zu untersuchende Metho-

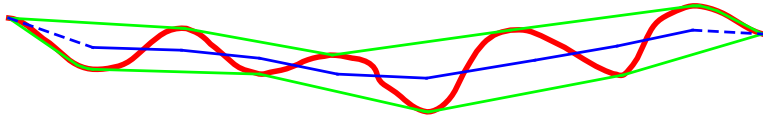


Abb. 5.2: Skizze eines Verfahrens zur Bestimmung eines Laufweges in Hauptbewegungsrichtung; die rote Trajektorie ist der originale Weg des verfolgten Kopfes; das grüne Polygon ist die Einhüllende zeitlich begrenzter konvexer Hüllen der detektierten Positionen eines Kopfes; der blaue Weg verläuft im gleichen Abstand zum oberen und unteren Segment der Einhüllenden und stellt den geglätteten Laufweg dar, dabei müssten die gestrichelten Enden gesondert behandelt und könnte durch Interpolation der Weg von abrupten Richtungsänderungen befreit werden

de könnte die Frequenzanalyse darstellen, die über die Positionsdaten oder abgeleiteten mikroskopischen Geschwindigkeits- oder Beschleunigungsdaten neben der Schrittfrequenz als Vorverarbeitung insbesondere niederfrequente Schwankungen möglicherweise aufspüren und glätten kann. Die Berücksichtigung der Kopfausrichtung bei Marker 1 oder 3 könnte ebenfalls die Bestimmung der Hauptbewegungsrichtung unterstützen. Der Einschluss des Laufweges in die Einhüllende bzw. Hüllkurve (*envelope*) zeitlich begrenzt an die Geschwindigkeit angepasster konvexer Hüllen der Kopfpositionen könnte ebenso über den Mittelweg eine Glättung des Laufweges ermöglichen (siehe Abb. 5.2).

Die Detektion und Verfolgung von Objekten besitzt viele Anwendungen über die der Extraktion von Laufwegen von Personen hinaus. Für viele Bereiche müsste zur Ermittlung von Trajektorien im Verfahrensablauf nur der Detektionsschritt durch eine dem Objekt angemessene Erkennung ersetzt werden, so wie es für die verschiedenen Marker bereits gemacht wurde. Die Fußgängerdynamik fällt in

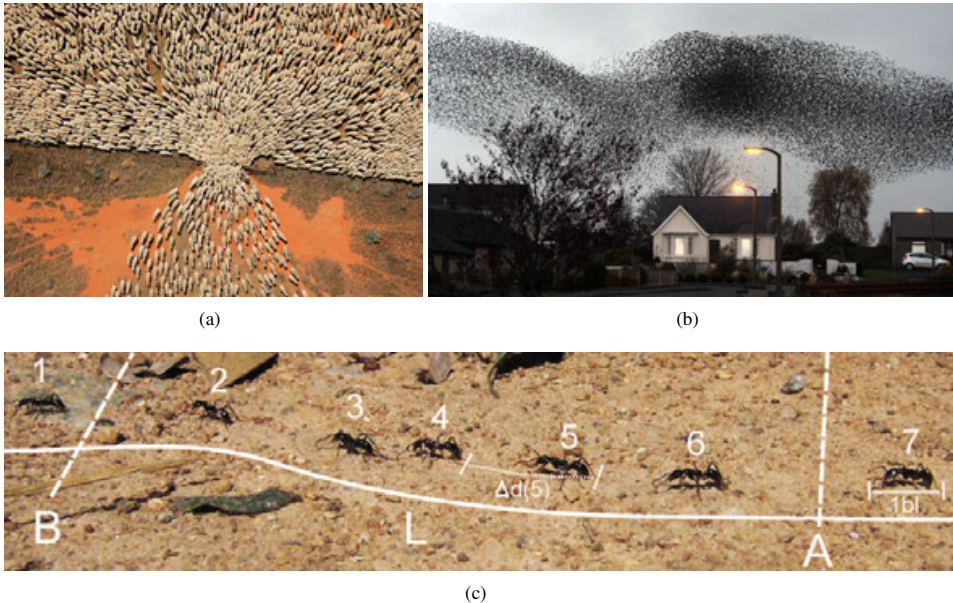


Abb. 5.3: Tiere als selbstgetriebene Teilchen: (a) Schafherde durch ein Viehgatter (wissenschaftlich behandelt in [620]), (b) Vogelschwarm und (c) Ameisenstraße; Bild aus (a) [621], (c) [622]

den Bereich der Dynamik selbstgetriebener Teilchen (*self-driven particles*). Hierzu gehört u. a. die Bewegung von Tieren (siehe Abb. 5.3). Um Gemeinsamkeiten von Systemen, wie z. B. der Stauform direkt vor der Engstelle in Abb. 5.3 a, selbstgetriebener Teilchen untersuchen und ggf. Theorien und Modelle adaptieren zu können, könnten die entwickelten Verfahren für Wissenschaftler verwandter Disziplinen nutzbar gemacht werden. Für Vögel [623, 624] und Fische [625] sind bereits Verfahren zur individuellen Verfolgung entwickelt worden. In [626] wird eine Software vorgestellt, die biologische Systeme auf kleiner Skala untersuchen kann. Dies wird an Experimenten mit Schaben gezeigt. Ebenfalls ein breiterer Ansatz wird in [227] verfolgt, wo Bereiche gleicher Bewegungsrichtung gesucht und zu einem einzelnen Objekt gruppiert werden. Aufgeführte Beispiele verfolgter Objekte sind neben Menschen und Autos auch Ameisen. Durch die Ähnlichkeit von Tieren gleicher Rasse bzw. in der untersten Ebene der biologischen Systematik ist zu erwarten, dass durch einen einfachen Vorlagen-Abgleich (siehe Abschnitt 2.3.1.3) bereits gute Ergebnisse erzielt werden.

Die Rechenzeit zur Bestimmung der Laufwege war bisher von untergeordnetem Interesse (siehe Abschnitt 3.6.8), da zum einen im gesamten Prozess bis zur Analyse der Trajektorien die automatische Extraktion der Laufwege nur einen relativ geringen zeitlichen Raum einnimmt und zum anderen für eine manuelle Sichtung bzw. qualitative Analyse der Versuche, die originalen Videoaufnahmen erhalten bleiben sollten. Daher war eine nachgelagerte Ermittlung der Trajektorien möglich. Dies könnte sich bei Feldstudien ändern, da zum einen durch die Länge der Beobachtungszeiträume das Datenvolumen so stark ansteigt, dass eine Speicherung einen zu großen Aufwand bedürfte (z. B. 1300 GB an Rohdaten einer Stereokamera bei der Feldstudie aus Abschnitt 3.6.7.2), und zum anderen die Persönlichkeitsrechte (Recht am eigenen Bild) [516] gewahrt bleiben müssen. Zudem würde ein Echtzeit- bzw. Online-Verfahren es ermöglichen, seine Versuchsplanung an die ermittelten Daten anzupassen, also z. B. bei der Variation der Dichte den besonders interessanten Phasenübergang zwischen Stehen und Gehen genauer abzutasten. Tab. 3.10 zeigt abgesehen von Extrembeispielen bereits eine Bildwiederholrate von 1,2 f/s für die markerlose Erkennung und bis zu 5,5 f/s bei der Detektion mit Markern bei Nutzung eines Threads auf einem handelsüblichen Laptop. In Abschnitt 3.6.8 sind bereits einige Ansätze zur Beschleunigung durch Optimierung oder Parallelisierung genannt worden. Die erforderliche Bildwiederholrate für ein Echtzeitsystem zur Extraktion von Laufwegen muss nicht mit der Rate des erfassenden Systems übereinstimmen, da die Verfolgung bei normalen Personengeschwindigkeiten bereits bei ca. 10 f/s zuverlässig funktionieren sollte (siehe Abschnitt 4.1.1.5). Eine Portierung der Laufwegerechnung auf eine intelligente Kamera (*smart camera*) würde darüber hinaus wegen der nicht mehr zu transportierenden Bilddaten die Übertragung vereinfachen, da die Datenmenge der Positionsdaten einen wesentlich geringeren Umfang besitzt.

Es ist in Veröffentlichungen zu beobachten, dass zum Teil die Auswertung von Experimenten oder Feldstudien mit einem unnötig großen Fehler durchgeführt werden, da für gewöhnlich die Autoren keine Experten im Bereich der Bildanalyse sind, sondern die Daten benötigen, weil sie die Laufwege analysieren oder ihre Modelle kalibrieren oder validieren möchten. Der häufigste Fehler ist die Vernachlässigung der optischen oder perspektivischen Verzerrung. Auch wenn die Software *PeTrack* frei zur Verfügung gestellt wird, setzt die Bedienung ein Verständnis vom Abbildungs- und Extraktionsprozess voraus. Um anderen die Auswertung von Laborversuchen im Bereich der Fußgängerdyamik zu erleichtern, führen wir daher im Oktober 2014 einen Workshop zu diesem Thema durch [627].

Dies waren eine Vielzahl von möglichen zukünftigen Aktivitäten oder Anpassungen der Verfahren, die zum Teil bereits begonnen wurden (Schrägsicht, Kinect, Benutzerunterstützung) und deren Umsetzung sich ansonsten nach dem Bedarf richten wird.

Literaturverzeichnis

- [1] Mulligan, Gordon F. und Jason P. Crampton: *Population growth in the world's largest cities*. Cities, 22(5):365 – 380, 2005. (Zitiert auf Seite 1)
- [2] Un-Habitat und United Nations Human Settlements Programme: *State of the World's Cities 2008/2009: Harmonious Cities*. The state of the world's cities report. Earthscan, 2008. (Zitiert auf Seite 1)
- [3] Wikipedia: *Haddsch*. <http://de.wikipedia.org/wiki/Haddsch> (12.02.2013). (Zitiert auf Seite 2)
- [4] Wikipedia: *Kumbh Mela*. http://de.wikipedia.org/wiki/Kumbh_Mela (12.02.2013). (Zitiert auf Seite 2)
- [5] ARGEBAU Fachkommission Bauaufsicht: *Musterverordnung über den Bau und Betrieb von Versammlungsstätten (Muster-Versammlungsstättenverordnung - MVStättV)*, 2005. (Zitiert auf den Seiten 2 und 208)
- [6] ARGEBAU Fachkommission Bauaufsicht: *MVStättV - Begründung und Erläuterungen: Musterverordnung über den Bau und Betrieb von Versammlungsstätten*, 2005. www.is-argebau.de. (Zitiert auf den Seiten 2 und 208)
- [7] Predtechenskii, Vsevolod Mikhailovich und Anatolii Ivanovich Milinskii: *Planning for Foot Traffic Flow in Buildings*. Amerind Publishing, New Dehli, 1978. Translation of: Proekttirovanie Zhdanii s Uchetom Organizatsii Dvizheniya Lyuddskikh Potokov, Stroiizdat Publishers, Moscow, 1969. (Zitiert auf Seite 2)
- [8] Weidmann, Ulrich: *Transporttechnik der Fussgänger*. Technical Report Schriftenreihe des IVT Nr. 90, Institut für Verkehrsplanung, Transporttechnik, Strassen- und Eisenbahnbau, ETH Zürich, 1993. (Zitiert auf Seite 2)
- [9] Nelson, Harold E. und Frederick W. Mowrer: *Emergency Movement*. In: DiNunno, P. J. (Herausgeber): *SFPE Handbook of Fire Protection Engineering*, Kapitel 14, Seite 367–380. National Fire Protection Association, Quincy MA, 2002. (Zitiert auf Seite 2)
- [10] Thompson, Peter A. und Eric W. Marchant: *A computer model for the evacuation of large building populations*. Fire Safety Journal, 24(2):131 – 148, 1995. (Zitiert auf Seite 2)
- [11] TraffGo HT GmbH: *Handbuch PedGo 2, PedGo Editor 2*, 2005. <http://www.evacuation-simulation.com/> (13.07.2014). (Zitiert auf Seite 2)

- [12] Kretz, Tobias, Stefan Hengst und Peter Vortisch: *Pedestrian Flow at Bottlenecks - Validation and Calibration of Vissim's Social Force Model of Pedestrian Traffic and its Empirical Foundations*. In: Sarvi, M (Herausgeber): *International Symposium of Transport Simulation 2008 (ISTS08)*, 2008. (Zitiert auf den Seiten 2 und 15)
- [13] Hostikka, Simo, Timo Korhonen, Tuomas Paloposki, Tuomo Rinne, Katri Matikainen und Simo Heliövaara: *Development and validation of FDS+Evac for evacuation simulations*. Technical Report, VTT Technical Research Centre of Finland, 2008. (Zitiert auf den Seiten 2, 14, 16, 42 und 43)
- [14] Rogsch, Christian: *Vergleichende Untersuchungen zur dynamischen Simulation von Personenströmen*. Diplomthesis, Bergische Universität Wuppertal, 2005. (Zitiert auf Seite 2)
- [15] Rogsch, Christian, Armin Seyfried und Wolfram Klingsch: *Comparative Investigations of the Dynamical Simulation of Foot Traffic Flow*. In: WALDAU, N. et al. [628], Seite 357–362. (Zitiert auf Seite 3)
- [16] Rogsch, Christian, Wolfram Klingsch, Armin Seyfried und Henning Weigel: *How reliable are commercial softwaretools for evacuation simulation?* In: *Interflam 2007* [629], Seite 235–246. (Zitiert auf Seite 3)
- [17] Schadschneider, Andreas, Wolfram Klingsch, Hubert Kluepfel, Tobias Kretz, Christian Rogsch und Armin Seyfried: *Encyclopedia of Complexity and System Science*, Band 5, Kapitel Evacuation Dynamics: Empirical Results, Modeling and Applications, Seite 3142–3176. Springer Berlin Heidelberg, 2009. (Zitiert auf den Seiten 3 und 206)
- [18] Seyfried, Armin, Oliver Passon, Bernhard Steffen, Maik Boltes, Tobias Rupprecht und Wolfram Klingsch: *New insights into pedestrian flow through bottlenecks*. *Transportation Science*, 43(3):395–406, 2009. (Zitiert auf den Seiten 3 und 206)
- [19] Liddle, Jack, Armin Seyfried, Bernhard Steffen, Wolfram Klingsch, Tobias Rupprecht, Andreas Winkens und Maik Boltes: *Microscopic insights into pedestrian motion through a bottleneck, resolving spatial and temporal variations*. arXiv e-prints, 2011. <http://arxiv.org/abs/1105.1532> (10.07.2014). (Zitiert auf den Seiten 3, 178, 206, 209, 210 und 211)
- [20] Seyfried, Armin, Andrea Portz und Andreas Schadschneider: *Phase Coexistence in Congested States of Pedestrian Dynamics*. In: Bandini, Stefania, Sara Manzoni, Hiroshi Umeo und Giuseppe Vizzari (Hrsg.): *Cellular Automata*, Band 6350 der Reihe *Lecture Notes in Computer Science*, Seite 496–505, 2010. (Zitiert auf den Seiten 3 und 206)
- [21] Zhang, Jun, Wolfram Klingsch, Andreas Schadschneider und Armin Seyfried: *Ordering in bi-directional pedestrian flows and its influence on the fundamental diagram*. *Journal of Statistical Mechanics: Theory and Experiment*, P02002(02):P02002, 2012. (Zitiert auf den Seiten 3, 206 und 207)
- [22] Nguyen, T.: *Analyse der Grundlagen des Fundamentaldiagramms für die Fußgängerbewegung*. Diplomthesis, Bergische Universität Wuppertal, 2006. (Zitiert auf Seite 3)
- [23] Sun, Deqing, Stefan Roth und Michael J. Black: *Secrets of optical flow estimation and their principles*. In: *Conference on Computer Vision and Pattern Recognition*, Seite 2432–2439, 2010. (Zitiert auf den Seiten 7 und 18)

- [24] Weickert, Joachim, Andrés Bruhn, Thomas Brox und Nils Papenberg: *A survey on variational optic flow methods for small displacements*. In: *Mathematical models for registration and applications to medical imaging*, Seite 103–136. Springer, 2006. (Zitiert auf Seite 7)
- [25] Brune, Christoph: *Berechnung des Optischen Flusses und Kantenerkennung mit Optimierungsmethoden*. Diplomarbeit, Westfälische Wilhelms-Universität Münster, 2007. (Zitiert auf Seite 7)
- [26] Black, Michael J. und P. Anandan: *A framework for the robust estimation of optical flow*. In: *Proceeding of IEEE International Conference on Computer Vision (ICCV)*, Seite 231–236, 1993. (Zitiert auf Seite 8)
- [27] Ellis, Andrew W., Andrew W. Young et al.: *Human cognitive neuropsychology: A textbook with readings*. Psychology Press, 2013. (Zitiert auf Seite 8)
- [28] Biederman, Irving: *Recognition-by-components: a theory of human image understanding*. *Psychological Review*, 94(2):115–147, 1987. (Zitiert auf Seite 8)
- [29] Eysenck, Michael W.: *Cognitive psychology: A student's handbook*. Taylor & Francis, 2000. (Zitiert auf Seite 8)
- [30] Nischwitz, Alfred, Max Fischer, Peter Haberäcker und Gudrun Socher: *Computergrafik und Bildverarbeitung: alles für Studium und Praxis - Bildverarbeitungswerkzeuge, Beispiel-Software und interaktive Vorlesungen online verfügbar*. Vieweg, 2007. (Zitiert auf den Seiten 8 und 171)
- [31] Boltes, Maik, Jun Zhang, Armin Seyfried und Bernhard Steffen: *T-junction: Experiments, trajectory collection, and analysis*. In: *IEEE International Conference on Computer Vision Workshop on Modeling, Simulation and Visual Analysis of Large Crowds [630]*, Seite 158–165. (Zitiert auf den Seiten 8, 206, 207, 212 und 214)
- [32] Boltes, Maik und Armin Seyfried: *Tracking People in Crowded Scenes*. In: WEIDMANN, ULRICH et al. [631], Seite 533–542. (Zitiert auf den Seiten 8 und 206)
- [33] Teixeira, Thiago, Gershon Dublon und Andreas Savvides: *A survey of human-sensing: Methods for detecting presence, count, location, track, and identity*. *ACM Computing Surveys*, 5, 2010. (Zitiert auf den Seiten 9, 14, 15, 41, 42, 45 und 179)
- [34] Jafari, Omid Hosseini, Dennis Mitzel und Bastian Leibe: *Real-Time RGB-D based People Detection and Tracking for Mobile Robots and Head-Worn Cameras*. In: *IEEE International Conference on Robotics and Automation (ICRA)*, 2014. (Zitiert auf den Seiten 9, 28, 29, 30 und 37)
- [35] Keller, Christoph Gustav: *Stereo-based Pedestrian Detection and Path Prediction*. Dissertation, Universität Heidelberg, 2014. (Zitiert auf Seite 9)
- [36] Borstell, Hagen, Bernd Gebert, Liu Cao, Cathrin Plate und Klaus Richter: *Image-Based Situation Assessment in Public Space*. In: Aschenbruck, Nils, Peter Martini, Michael Meier und Jens Tölle (Hrsg.): *Future Security*, Band 318 der Reihe *Communications in Computer and Information Science*, Seite 61–64. Springer Berlin Heidelberg, 2012. (Zitiert auf Seite 9)
- [37] Raumfahrt, Deutsches Zentrum für Luft- und Raumfahrt: *VABENE++ - Verkehrsmanagement bei Großereignissen und Katastrophen*. <http://www.dlr.de/vabene/> (12.06.2014). (Zitiert auf Seite 9)

- [38] Wu, Shandong, Brian E. Moore und Mubarak Shah: *Chaotic Invariants of Lagrangian Particle Trajectories for Anomaly Detection in Crowded Scenes*. In: *IEEE Conference on Computer Vision and Pattern Recognition*, 2010. (Zitiert auf den Seiten 9, 11 und 23)
- [39] Velastin, Sergio A., Boghos A. Boghossian und Maria Alicia Vicencio-Silva: *A motion-based image processing system for detecting potentially dangerous situations in underground railway stations*. *Transportation Research Part C: Emerging Technologies*, 14(2):96–113, 2006. (Zitiert auf Seite 9)
- [40] Meyer, Kenneth, Hugh L. Applewhite und Frank A. Biocca: *A survey of position trackers*. In: *Presence: Teleoperators and Virtual Environments*, Band 1, Seite 173–200, 1992. (Zitiert auf Seite 9)
- [41] Bedagkar-Gala, Apurva und Shishir K. Shah: *A Survey of Approaches and Trends in Person Re-identification*. *Image and Vision Computing*, 2014. (Zitiert auf Seite 9)
- [42] Mazzon, Riccardo, Syed Fahad Tahir und Andrea Cavallaro: *Person re-identification in crowd*. *Pattern Recognition Letters*, 33(14):1828–1837, 2012. (Zitiert auf Seite 9)
- [43] Zhao, Rui, Wanli Ouyang und Xiaogang Wang: *Person Re-identification by Saliency Learning*. *arXiv e-prints*, 2014. (Zitiert auf Seite 9)
- [44] Hjeltnäs, Erik und Boon Kee Low: *Face Detection: A Survey*. *Computer Vision and Image Understanding*, 83(3):236 – 274, 2001. (Zitiert auf Seite 9)
- [45] Delac, Kresimir und Mislav Grgic (Hrsg.): *Face Recognition*. I-Tech Education and Publishing, 2007. (Zitiert auf Seite 9)
- [46] Hua, Gang, Ming-Hsuan Yang, E. Learned-Miller, Yi Ma, M. Turk, D. Kriegman und T.S. Huang: *Introduction to the Special Section on Real-World Face Recognition*. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(10):1921–1924, 2011. (Zitiert auf Seite 9)
- [47] Jean, Frederic, Robert Bergevin und Alexandra Branzan Albu: *Human Gait Characteristics from Unconstrained Walks and Viewpoints*. In: *Proceeding of IEEE International Conference on Computer Vision (ICCV)* [630], Seite 1883–1888. (Zitiert auf Seite 9)
- [48] Tistarelli, Massimo, Josef Bigun und Anil K. Jain (Hrsg.): *Biometric Authentication*, Band 2359 der Reihe *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2002. (Zitiert auf Seite 9)
- [49] Awad, Fahed und Rufaida Shamroukh: *Human Detection by Robotic Urban Search and Rescue Using Image Processing and Neural Networks*. *International Journal of Intelligence Science*, 4:39–53, 2014. (Zitiert auf den Seiten 9, 20 und 35)
- [50] Wiskott, Laurenz, Jean-Marc Fellous, Norbert Krüger und Christoph von der Malsburg: *Face recognition and gender determination*. In: *Int. Workshop on Automatic Face- and Gesture-Recognition*, Seite 92–97, 1995. (Zitiert auf Seite 9)
- [51] Zhao, W., R. Chellappa, P. J. Phillips und A. Rosenfeld: *Face Recognition: A Literature Survey*. *ACM Comput. Surv.*, 35(4):399–458, 2003. (Zitiert auf Seite 9)

- [52] Collins, Robert T, Alan Lipton, Takeo Kanade, Hironobu Fujiyoshi, David Duggins, Yanghai Tsin, David Tolliver, Nobuyoshi Enomoto, Osamu Hasegawa, Peter Burt et al.: *A system for video surveillance and monitoring*, Band 2. Carnegie Mellon University, the Robotics Institute Pittsburg, 2000. (Zitiert auf den Seiten 9 und 11)
- [53] Siebel, N.T.: *Design and Implementation of People Tracking Algorithms for Visual Surveillance Applications*. Dissertation, The University of Reading, 2003. (Zitiert auf Seite 9)
- [54] Konrad, Janusz (Herausgeber): *10th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Kraków, Poland, 2013. Polish Society for Theoretical and Applied Electrical Engineering (PTETiS). (Zitiert auf Seite 9)
- [55] Chraibi, Mohcine, Armin Seyfried und Andreas Schadschneider: *Generalized centrifugal force model for pedestrian dynamics*. Physical Review E, 82:046111, 2010. (Zitiert auf den Seiten 11, 127, 206, 215 und 284)
- [56] Boltes, Maik und Armin Seyfried: *Collecting Pedestrian Trajectories*. Neurocomputing, Special Issue on Behaviours in Video, 100:127–133, 2013. (Zitiert auf den Seiten 11 und 206)
- [57] Boltes, Maik, Jun Zhang und Armin Seyfried: *Analysis of crowd dynamics with laboratory experiments*. In: ALI, SAAD et al. [112], Kapitel 4, Seite 67 – 98. ISBN: 978-1-4614-8482-0 (Print) 978-1-4614-8483-7 (Online). (Zitiert auf den Seiten 11, 206, 208, 209 und 212)
- [58] Chraibi, Mohcine, Andreas Schadschneider und Armin Seyfried: *On Force-Based Modeling of Pedestrian Dynamics*. In: ALI, SAAD et al. [112], Seite 23 – 41. ISBN: 978-1-4614-8482-0 (Print) 978-1-4614-8483-7 (Online). (Zitiert auf Seite 11)
- [59] Krausz, Barbara und Christian Bauckhage: *Loveparade 2010: Automatic video analysis of a crowd disaster*. Computer Vision and Image Understanding, Special issue on Semantic Understanding of Human Behaviors in Image Sequences, 116(3):307 – 319, 2012. (Zitiert auf den Seiten 10 und 11)
- [60] Ali, Saad und Mubarak Shah: *A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis*. In: *Conference on Computer Vision and Pattern Recognition*, 2007. (Zitiert auf den Seiten 10, 11 und 23)
- [61] Muybridge, Eadweard: *Animal locomotion, Plate 519*. Da Capo Press New York, 1887. (Zitiert auf Seite 10)
- [62] Becker, Markus: *Simulierter Patient: Operieren am virtuellen Menschen*. Spiegel Online Wissenschaft, Seite 1, 2014. (Zitiert auf den Seiten 10 und 11)
- [63] Cristani, Marco, R. Raghavendra, Alessio Del Bue und Vittorio Murino: *Human behavior analysis in video surveillance: A Social Signal Processing perspective*. Neurocomputing, 100:86–97, 2013. (Zitiert auf Seite 11)
- [64] Zhao, Tao und Ram Nevatia: *Tracking Multiple Humans in Complex Situations*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(9):1208–1221, 2004. (Zitiert auf Seite 11)

- [65] Brostow, Gabriel J. und Roberto Cipolla: *Unsupervised Bayesian Detection of Independent Motion in Crowds*. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2006, Vol. 1*, Band I, Seite 594–601, 2006. (Zitiert auf den Seiten 11, 22 und 23)
- [66] Stalder, S., H. Grabner und L. Van Gool: *Exploring Context to Learn Scene Specific Object Detectors*. In: *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Miami, USA*, Seite 63–70, 2009. (Zitiert auf Seite 11)
- [67] Antonini, Gianluca, Santiago Venegas Martinez, Michel Bierlaire und Jean Philippe Thiran: *Behavioral priors for detection and tracking of pedestrians in video sequences*. *International Journal of Computer Vision*, 69(2):159–180, 2006. (Zitiert auf Seite 11)
- [68] Rodriguez, Mikel, Saad Ali und Takeo Kanade: *Tracking in unstructured crowded scenes*. In: *Proceeding of IEEE International Conference on Computer Vision (ICCV)*, Seite 1389–1396, 2009. (Zitiert auf den Seiten 11, 23, 45 und 70)
- [69] Johansson, Anders Fredrik: *Data-Driven Modeling of Pedestrian Crowds*. Dissertation, Faculty of Traffic Sciences Friedrich List Technische Universität Dresden, 2009. (Zitiert auf den Seiten 11 und 23)
- [70] Johansson, Anders, Dirk Helbing, Habib Z. Al-Abideen und Salim Al-Bosta: *From Crowd Dynamics to Crowd Safety: A Video-Based Analysis*. *Advances in Complex Systems (ACS)*, 11:4, 2008. (Zitiert auf den Seiten 11, 22 und 23)
- [71] Cui, Jing, Weibin Liu und Weiwei Xing: *Crowd behaviors analysis and abnormal detection based on surveillance data*. *Journal of Visual Languages & Computing*, 25(6):628 – 636, 2014. *Distributed Multimedia Systems {DMS2014} Part I*. (Zitiert auf Seite 11)
- [72] Pathan, Saira Saleem und Klaus Richter: *Pedestrian Behavior Analysis with Image-Based Method in Crowds*. In: *Traffic and Granular Flow 2013*, 2015. (Zitiert auf den Seiten 11 und 23)
- [73] Boghossian, A. und A. Velastin: *Motion-based machine vision techniques for the management of large crowds*. In: *Proceedings of 6th IEEE International Conference on ICECS Electronics, Circuits and Systems*, Band 2, Seite 961–964, 2002. (Zitiert auf Seite 11)
- [74] Andrade, L., S. Blunsden und R. Fisher: *Modeling Crowd Scenes for Event Detection*. In: *Proceedings of the 18th International Conference on Pattern Recognition*, Seite 175–178, 2006. (Zitiert auf Seite 11)
- [75] Andrade, Ernesto L., Blunsden Scott und Robert B. Fisher: *Hidden Markov Models for Optical Flow Analysis in Crowds*. In: *Proceedings of the 18th International Conference on Pattern Recognition*, Seite 460–463, 2006. (Zitiert auf Seite 11)
- [76] Kratz, Louis und Ko Nishino: *Anomaly Detection in Extremely Crowded Scenes Using Spatio-Temporal Motion Pattern Models*. In: *IEEE Conference on Computer Vision and Pattern Recognition*, Seite 1446 – 1453, 2009. (Zitiert auf Seite 11)
- [77] Mehran, Ramin, Alexis Oyama und Mubarak Shah: *Abnormal crowd behavior detection using social force model*. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 0:935–942, 2009. (Zitiert auf den Seiten 11 und 23)

- [78] Albiol, A., M. Silla, A. Albiol und J. Mossi: *Video Analysis using Corner Motion Statistics*. In: *Performance Evaluation of Tracking and Surveillance workshop at CVPR 2009*, Seite 31–37, 2009. (Zitiert auf Seite 11)
- [79] Benabbas, Y., N. Ihaddadene und C. Djeraba: *Global analysis of motion vectors for event detection in crowd scenes*. In: *Performance Evaluation of Tracking and Surveillance workshop at CVPR 2009*, Seite 109–116, 2009. (Zitiert auf Seite 11)
- [80] Chan, A. B., M. Morrow und N. Vasconcelos: *Analysis of Crowded scenes using Holistic Properties*. In: *Performance Evaluation of Tracking and Surveillance workshop at CVPR*, Seite 31–37, 2009. (Zitiert auf Seite 11)
- [81] Mahadevan, V., W.X. Li, V. Bhalodia und N. Vasconcelos: *Anomaly detection in crowded scenes*. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Seite 1975–1981, 2010. (Zitiert auf Seite 11)
- [82] Helbing, Dirk, Anders Johansson und Habib Zein Al-Abideen: *Dynamics of Crowd Disasters: An Empirical Study*. *Physical Review E*, 75:046109, 2007. (Zitiert auf Seite 11)
- [83] Qu, Jianming, Zhijing Liu und Wenhua He: *Video Pedestrian Detection Based on Orthogonal Scene Motion Pattern*. *Mathematical Problems in Engineering*, 2014:8, 2014. (Zitiert auf Seite 11)
- [84] Lopez, M., SA Velastin und M. Rodriguez: *Passenger density measurement in a train carriage using image processing*. *Transportation systems*, Seite 961–965, 1997. (Zitiert auf Seite 11)
- [85] Ryan, David, Simon Denman, Sridha Sridharan und Clinton Fookes: *An Evaluation of Crowd Counting Methods, Features and Regression Models*. *Computer Vision and Image Understanding*, 2014. (Zitiert auf Seite 11)
- [86] Chen, Ke, Chen Change Loy, Shaogang Gong und Tao Xiang: *Feature mining for localised crowd counting*. In: *Proceedings of British Machine Vision Conference (MBVC)*, 2012. (Zitiert auf Seite 11)
- [87] Lempitsky, V. und A. Zisserman: *Learning To Count Objects in Images*. In: *Advances in Neural Information Processing Systems*, 2010. (Zitiert auf Seite 11)
- [88] Abuarafah, Adnan Ghazi, Mohamed Osama Khozium und Essam AbdRabou: *Real-time Crowd Monitoring using Infrared Thermal Video Sequences*. *Journal of American Science*, 8(3):133–140, 2012. (Zitiert auf Seite 11)
- [89] Ismail, K., T. Sayed und N. Saunier: *Automated Observation of Pedestrian Walking Speed Using Computer Vision Techniques*. *Journal of Intelligent Transportation Systems Technology Planning and Operations*, 2011. (Zitiert auf den Seiten 11 und 15)
- [90] Wikipedia: *Eadweard Muybridge*. http://de.wikipedia.org/wiki/Eadweard_Muybridge (12.05.2014). (Zitiert auf Seite 11)
- [91] Gmiterko, Alexander und Tomas Liptak: *Motion Capture of Human for Interaction with Service Robot*. *American Journal of Mechanical Engineering*, 1(7):212–216, 2013. (Zitiert auf Seite 11)

- [92] Wikipedia: *Kinect*. <http://de.wikipedia.org/wiki/Kinect> (12.05.2014). (Zitiert auf den Seiten 11 und 37)
- [93] Ye, Mao, Xianwang Wang, Ruigang Yang, Liu Ren und Marc Pollefeys: *Accurate 3D Pose Estimation From a Single Depth Image*. In: *Proceeding of IEEE International Conference on Computer Vision (ICCV)* [630], Seite 731–738. (Zitiert auf Seite 11)
- [94] Baak, Andreas, Meinard Müller, Gaurav Bharaj, Hans-Peter Seidel und Christian Theobalt: *A Data-Driven Approach for Real-Time Full Body Pose Reconstruction from a Depth Camera*. In: *Proceeding of IEEE International Conference on Computer Vision (ICCV)* [630], Seite 1092–1099. (Zitiert auf Seite 11)
- [95] Bauer, Sebastian, Jakob Wasza, Sven Haase, Natalia Marosi und Joachim Hornegger: *Multi-modal Surface Registration for Markerless Initial Patient Setup in Radiation Therapy using Microsoft's Kinect Sensor*. In: *Proceeding of IEEE International Conference on Computer Vision (ICCV)* [630], Seite 1175–1181. (Zitiert auf den Seiten 11 und 37)
- [96] Gavrilu, D.M: *The Visual Analysis of Human Movement: A Survey*. Computer Vision and Image Understanding, 73(1):82 – 98, 1999. (Zitiert auf Seite 13)
- [97] Moeslund, T.B. und E. Granum: *A survey of computer vision-based human motion capture*. Computer Vision and Image Understanding, 81(3):231–268, 2001. (Zitiert auf Seite 13)
- [98] Yilmaz, Alper, Omar Javed und Mubarak Shah: *Object Tracking: A Survey*. ACM Comput. Surv., 38(4), 2006. (Zitiert auf Seite 13)
- [99] Ogale, Neeti A: *A survey of techniques for human detection from video*. Diplomarbeit, University of Maryland, 2006. (Zitiert auf den Seiten 13 und 179)
- [100] Zhan, Beibei, Dorothy Ndedi Monekosso, Paolo Remagnino, Sergio A. Velastin und Li-Qun Xu: *Crowd analysis: a survey*. Machine Vision Applications, 19(5-6):345–357, 2008. (Zitiert auf den Seiten 13 und 23)
- [101] Roth, Peter M. und Martin Winter: *Survey of Appearance-Based Methods for Object Recognition*. Technical Report, Institute for Computer Graphics and Vision, Graz University of Technology, 2008. (Zitiert auf Seite 13)
- [102] Nikos, Nikolaidis, Pantic Maja und Pitas Ioannis (Hrsg.): *Anthropocentric Video Analysis: Tools and Applications*, Band 2008 der Reihe *EURASIP Journal on Image and Video Processing*. Hindawi Publishing Corporation, 2008. (Zitiert auf den Seiten 13 und 278)
- [103] Kelly, Philip: *Pedestrian detection and tracking using stereo vision techniques*. Dissertation, Dublin City University, 2008. (Zitiert auf Seite 13)
- [104] Timmermans, H. J. P.: *Pedestrian Behavior: Models, Data Collection and Applications*. Emerald Group Pub Limited, 2009. (Zitiert auf den Seiten 13 und 278)
- [105] Schiele, Bernt, Mykhaylo Andriluka, Nikodem Majer, Stefan Roth und Christian Wojek: *Visual people detection - different models, comparison and discussion*. In: Arras, K.O. (Herausgeber): *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) 2009 workshop on people detection and tracking*, Seite 1–8, 2009. (Zitiert auf den Seiten 13 und 179)

- [106] Enzweiler, M. und D.M. Gavrilă: *Monocular Pedestrian Detection: Survey and Experiments*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 31(12):2179–2195, 2009. (Zitiert auf Seite 14)
- [107] Jacques Junior, J. C. S., S. Raupp Musse und C. R. Jung: *Crowd Analysis Using Computer Vision Techniques*. Signal Processing Magazine, IEEE, 27(5):66–77, 2010. (Zitiert auf den Seiten 14 und 23)
- [108] Moeslund, Thomas B., Adrian Hilton, Volker Krüger und Leonid Sigal (Hrsg.): *Visual Analysis of Humans - Looking at People*. Springer, 2011. (Zitiert auf den Seiten 14, 19, 240 und 278)
- [109] Dollár, Piotr, Christian Wojek, Bernt Schiele und Pietro Perona: *Pedestrian detection: An evaluation of the state of the art*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 34(4):743–761, 2012. (Zitiert auf den Seiten 14, 22 und 23)
- [110] Nguyen, Duc Thanh: *Human detection from images and videos*. Dissertation, School of Computer Science and Software Engineering, University of Wollongong, 2012. (Zitiert auf Seite 14)
- [111] Zhou, Huiyu, Yuan Yuan, Eliza Y Du und Pingkun Yan (Hrsg.): *Special Issue: Behaviours in Video*, Band 100 der Reihe *Neurocomputing*. Elsevier, 2013. (Zitiert auf den Seiten 14 und 278)
- [112] Ali, Saad, Ko Nishino, Dinesh Manocha und Mubarak Shah (Hrsg.): *Modeling, Simulation and Visual Analysis of Crowds: A Multidisciplinary Perspective*, Band 11 der Reihe *The International Series in Video Computing*. Springer New York, 2013. ISBN: 978-1-4614-8482-0 (Print) 978-1-4614-8483-7 (Online). (Zitiert auf den Seiten 14, 23, 229, 274 und 278)
- [113] Thida, Myo, YokeLeng Yong, Pau Climent-Pérez, How-lung Eng und Paolo Remagnino: *A Literature Review on Video Analytics of Crowded Scenes*. In: Atrey, Pradeep K., Mohan S. Kankanhalli und Andrea Cavallaro (Hrsg.): *Intelligent Multimedia Surveillance*, Seite 17–36. Springer Berlin Heidelberg, 2013. (Zitiert auf den Seiten 14 und 23)
- [114] Calderara, Simone, Stefania Bandini und Rita Cucchiara (Hrsg.): *Special Issue on Pattern Recognition and Crowd Analysis*, Band 44 der Reihe *Pattern Recognition Letters*. Elsevier, 2014. (Zitiert auf den Seiten 14, 23 und 278)
- [115] Plaue, Matthias, Minjie Chen, Günter Bärwolff und Hartmut Schwandt: *Trajectory Extraction and Density Analysis of Intersecting Pedestrian Flows from Video Recordings*. In: Stilla, Uwe, Franz Rottensteiner, Helmut Mayer, Boris Jutzi und Matthias Butenuth (Hrsg.): *Photogrammetric Image Analysis*, Band 6952 der Reihe *Lecture Notes in Computer Science*, Seite 285–296. Springer Berlin Heidelberg, 2011. (Zitiert auf den Seiten 15 und 181)
- [116] Kretz, Tobias, Anna Grünebohm und Michael Schreckenberg: *Experimental study of pedestrian flow through a bottleneck*. Journal of Statistical Mechanics: Theory and Experiment, 10:P10014, 2006. (Zitiert auf den Seiten 15, 181 und 208)
- [117] Kretz, Tobias, Anna Grünebohm, Maike Kaufman, Florian Mazur und Michael Schreckenberg: *Experimental study of pedestrian counterflow in a corridor*. Journal of Statistical Mechanics: Theory and Experiment, 10:P10001, 2006. (Zitiert auf Seite 15)
- [118] Fang, Zhiming, Weiguo Song, Jun Zhang und Hao Wu: *Experiment and modeling of exit-selecting behaviors during a building evacuation*. Physica A: Statistical Mechanics and its Applications, 389:815–824, 2010. (Zitiert auf den Seiten 15 und 181)

- [119] Ma, Jian, Wei-Guo Song und Guang-Xuan Liao: *Multi-grid simulation of pedestrian counter flow with topological interaction*. Chinese Physics B, 19(12):128901–1 – 128901–9, 2010. (Zitiert auf den Seiten 15 und 181)
- [120] Lam, William H. K., Jodie Y. S. Lee und C. Y. Cheung: *A study of the bi-directional pedestrian flow characteristics at Hong Kong signalized crosswalk facilities*. Transportation, 29:169–192, 2002. (Zitiert auf den Seiten 15 und 181)
- [121] Yanagisawa, Daichi, Ayako Kimura, Akiyasu Tomoeda, Nishi Ryosuke, Yushi Suma, Kazumichi Ohtsuka und Katsuhiro Nishinari: *Introduction of Frictional and Turning Function for Pedestrian Outflow with an Obstacle*. Physical Review E, 80:036110–1 – 036110–14, 2009. (Zitiert auf den Seiten 15 und 181)
- [122] Nagai, R., M. Fukamachi und T. Nagatani: *Evacuation of crawlers and walkers from corridor through an exit*. Physica A, 367:449–460, 2006. (Zitiert auf den Seiten 15 und 181)
- [123] Isobe, M., T. Adachi und T. Nagatani: *Experiment and simulation of pedestrian counter flow*. Physica A, 336:638–650, 2004. (Zitiert auf den Seiten 15 und 181)
- [124] Muir, H. C., D. M. Bottomley und C. Marrison: *Effects of Motivation and Cabin Configuration on Emergency Aircraft Evacuation Behavior and Rates of Egress*. The International Journal of Aviation Psychology, 6:57–77, 1996. (Zitiert auf Seite 15)
- [125] Helbing, D., M. Isobe, T. Nagatani und K. Takimoto: *Lattice gas simulation of experimentally studied evacuation dynamics*. Physical Review E, 67:067101, 2003. (Zitiert auf Seite 15)
- [126] Lv, Wei, Xiaoge Wei und Weiguo Song: *Experimental Study on the Interaction Mechanism of Cross-Walking Pedestrians*. In: *Traffic and Granular Flow 2013*, 2015. (Zitiert auf Seite 15)
- [127] Bamberger, Johanna, Anna-Lena Geßler, Peter Heitzelmann, Sara Korn, Rene Kahlmeyer, Xue Hao Lu, Qi Hao Sang, Zhi Jie Wang, Guan Zong Yuan, Michael Gauß und Tobias Kretz: *Crowd Research at School: Crossing Flows*. In: *Traffic and Granular Flow 2013*, 2015. (Zitiert auf Seite 15)
- [128] Nikolic, Marija, Bilal Farooq und Michel Bierlaire: *Exploratory analysis of pedestrian flow characteristics in mobility hubs using trajectory data*. In: *Proceedings of the Swiss Transportation Research Conference*, 2013. (Zitiert auf Seite 15)
- [129] Tu, Peter H., Gianfranco Doretto, Nils O. Krahnstoeve, A. G. Amitha Perera, Frederick W. Wheeler, Xiaoming Liu, Jens Rittscher, Thomas B. Sebastian, Ting Yu und Kevin G. Harding: *An intelligent video framework for homeland protection*. In: *Defense and Security Symposium*, Seite 65620C–65620C, 2007. (Zitiert auf Seite 15)
- [130] Grabner, Helmut, Peter M Roth und Horst Bischof: *Is pedestrian detection really a hard task*. In: *Proc. IEEE Intern. Workshop on Performance Evaluation of Tracking and Surveillance*, Seite 1–8, 2007. (Zitiert auf Seite 15)
- [131] Cao, XB, Yan Wu Xu, D Chen und Hong Qiao: *Associated evolution of a support vector machine-based classifier for pedestrian detection*. Information Sciences, 179(8):1070–1077, 2009. (Zitiert auf Seite 15)

- [132] Kang, Hee-Gu und Daijin Kim: *Real-time multiple people tracking using competitive condensation*. Pattern Recognition, 38(7):1045–1058, 2005. (Zitiert auf Seite 15)
- [133] Hoogendoorn, S.P., W. Daamen und P.H.L. Bovy: *Extracting microscopic pedestrian characteristics from video data*. In: *TRB2003 Annual Meeting*, 2003. (Zitiert auf den Seiten 16, 24 und 25)
- [134] Hoogendoorn, S. P. und W. Daamen: *Pedestrian Behavior at Bottlenecks*. Transportation Science, 39(2):147–159, 2005. (Zitiert auf Seite 16)
- [135] Daamen, Winnie und S. P. Hoogendoorn: *Controlled experiments to derive walking behaviour*. European Journal of Transport and Infrastructure Research, 3(1):39–59, 2003. (Zitiert auf Seite 16)
- [136] Daamen, Winnie, Enne de Boer und Robert de Kloe: *Assessing the gap between public transport vehicles and platforms as a barrier for the disabled: Use of laboratory experiments*. Transportation Research Record: Journal of the Transportation Research Board, 2072(1):131–138, 2008. (Zitiert auf Seite 16)
- [137] Daamen, W., S.P. Hoogendoorn, R. L. Landman und H. van Wijngaarden: *Pedestrian behavior experiments in and around automated revolving doors*. In: *TRB 87th Annual Meeting, Compendium of papers*, Seite 1–15, 2008. (Zitiert auf den Seiten 16 und 25)
- [138] TU Delft, Fachbereich Verkehr und Planung: *Pedestrian Data Collection*. <http://www.citg.tudelft.nl/en/about-faculty/departments/transport-and-planning/traffic-management-and-traffic-flow-theory/dynamisch-verkeers-management/special-projects/pedestrians/data-collection/> (30.06.2014). (Zitiert auf den Seiten 16 und 25)
- [139] Daamen, Winnie und Serge Hoogendoorn: *Capacity of doors during evacuation conditions*. Procedia Engineering, 1st Conference on Evacuation Modeling and Management, 3(0):53 – 66, 2010. (Zitiert auf den Seiten 16, 24, 25, 34 und 182)
- [140] Liu, Xuan, Weiguo Song und Jun Zhang: *Extraction and quantitative analysis of microscopic evacuation characteristics based on digital image processing*. Physica A: Statistical Mechanics and its Applications, 388(13):2717–2726, 2009. (Zitiert auf den Seiten 16, 24, 25 und 181)
- [141] Song, Weiguo, Jun Zhang und Armin Seyfried: *Experimental Study of Pedestrian Flow in the Channel through Bottleneck*. In: PEACOCK, R.D. et al. [632], Seite 875–879. (Zitiert auf den Seiten 16, 24, 25, 206 und 207)
- [142] Tian, Wei, Weiguo Song, Jian Ma, Zhiming Fang, Armin Seyfried und Jack Liddle: *Experimental study of pedestrian behaviors in a corridor based on digital image processing*. Fire Safety Journal, 47(0):8 – 15, 2012. (Zitiert auf den Seiten 16, 25 und 206)
- [143] Moussaïd, M., D. Helbing, S. Garnier, A. Johansson, M. Combe und G. Theraulaz: *Experimental study of the behavioural mechanisms underlying self-organization in human crowds*. Proceedings of the Royal Society B: Biological Sciences, 276(1668):2755–2762, 2009. (Zitiert auf den Seiten 16, 24, 25, 33 und 181)
- [144] Saadat, Saman, Kardi Teknomo und Proceso Fernandez: *Automation of Tracking Trajectories in a Crowded Situation*. Fire Technology, Seite 1–18, 2010. (Zitiert auf den Seiten 16, 22, 24 und 25)

- [145] Saadat, S. und K. Teknomo: *Automation of Pedestrian Tracking in a Crowded Situation*. In: Peacock, Richard D., Erica D. Kuligowski und Jason D. Averill (Hrsg.): *Pedestrian and Evacuation Dynamics*, Seite 231–239. Springer US, 2011. (Zitiert auf Seite 16)
- [146] Wong, S. C., W. L. Leung, S. H. Chan, Wiliam H. K. Lam, Nelson H. C. Yung, C. Y. Liu und Peng Zhang: *Bidirectional Pedestrian Stream Model with Oblique Intersecting Angle*. *Journal of Transportation Engineering*, 136(3):234–242, 2010. (Zitiert auf den Seiten 16, 24, 25, 26 und 181)
- [147] Ren-Yong, Guo, Wong S. C., Xia Yin-Hua, Huang Hai-Jun, Lam William H. K. und Choi Keechoo: *Empirical Evidence for the Look-Ahead Behavior of Pedestrians in Bi-directional Flows*. *Chinese Physics Letters*, 29(6):068901, 2012. (Zitiert auf den Seiten 16, 26 und 181)
- [148] Bukáček, Marek, Pavel Hrabák und Milan Krbálek: *Experimental Analysis of Two-Dimensional Pedestrian Flow in front of the Bottleneck – Experimental Analysis of 2D Pedestrian Flow*. In: *Traffic and Granular Flow 2013*, 2015. (Zitiert auf den Seiten 16, 24 und 26)
- [149] Christensen, K. M., M. S. Sharifi, D. Stuart, A. Chen, Y. Kim und Y. Chen: *Overview of a large-scale controlled experiment on pedestrian walking behavior involving individuals with disabilities*. In: *93rd Annual Meeting of the Transportation Research Board*, 2014. (Zitiert auf den Seiten 16 und 26)
- [150] Stuart, D., K. M. Christensen, A. Chen, Y. Kim und Y. Chen: *Utilizing augmented reality technology for crowd pedestrian analysis involving individuals with disabilities*. In: *Proceedings of the ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, 2013. (Zitiert auf den Seiten 16, 24 und 26)
- [151] Chen, Anthony, Mohammad Sadra Sharifi, Daniel Stuart, Keith Christensen, Yong Seog Kim und YangQuan Chen: *Analysis of walking speeds involving individuals with disabilities in different walking environments*. In: *93rd Annual Meeting of the Transportation Research Board*, 2014. (Zitiert auf den Seiten 16 und 26)
- [152] Nescher, Thomas und Andreas Kunz: *Using Head Tracking Data for Robust Short Term Path Prediction of Human Locomotion*. In: Gavrilova, Marina L., C.J. Kenneth Tan und Arjan Kuijper (Hrsg.): *Transactions on Computational Science XVIII*, Band 7848 der Reihe *Lecture Notes in Computer Science*, Seite 172–191. Springer Berlin Heidelberg, 2013. (Zitiert auf den Seiten 16 und 26)
- [153] Karkin, Ilya, Vladimir Grachev, Andrey Skochilov und Vladimir Zverev: *FlowTech and EvaTech: Two Computer-Simulation Methods for Evacuation Calculation*. In: Klingsch, Wolfram W. F., Christian Rogsch, Andreas Schadschneider und Michael Schreckenberg (Hrsg.): *Pedestrian and Evacuation Dynamics 2008*, Seite 537–545, 2010. (Zitiert auf den Seiten 16 und 36)
- [154] Hostikka, Simo, Tuomas Paloposki, Tuomo Rinne, Juha Matti und Timo Korhonen: *Evacuation experiments in offices and public buildings*. VTT WORKING PAPERS, 2007. (Zitiert auf den Seiten 16, 43 und 44)
- [155] Lemerrier, Samuel, Mathieu Moreau, Mehdi Moussaïd, Guy Theraulaz, Stéphane Donikian und Julien Pettré: *Reconstructing Motion Capture Data for Human Crowd Study*. In: Allbeck, Jan M. und Petros Faloutsos (Hrsg.): *Motion in Games*, Band 7060 der Reihe *Lecture Notes in Computer Science*, Seite 365–376. Springer Berlin Heidelberg, 2011. (Zitiert auf den Seiten 16 und 34)

- [156] Jelić, Asja, Cécile Appert-Rolland, Samuel Lemerrier und Julien Pettré: *Properties of pedestrians walking in line: Stepping behavior*. Physical Review E, 86:046111, 2012. (Zitiert auf den Seiten 16, 34, 72, 127 und 181)
- [157] Moussaïd, Mehdi, Elsa G. Guilloit, Mathieu Moreau, Jerome Fehrenbach, Olivier Chabiron, Samuel Lemerrier, Julien Pettré, Cecile Appert-Rolland, Pierre Degond und Guy Theraulaz: *Traffic Instabilities in Self-Organized Pedestrian Crowds*. PLoS Computational Biology, 8:1002442, 2012. (Zitiert auf den Seiten 16, 34 und 181)
- [158] University College London: *Pedestrian Accessibility Movement Environment Laboratory (pamela)*. <http://www.ucl.ac.uk/arg/pamela> (12.07.2014). (Zitiert auf den Seiten 16, 40 und 41)
- [159] Bauer, Dietmar und Kay Kitazawa: *Using laser scanner data to calibrate certain aspects of microscopic pedestrian motion models*. In: KLINGSCH, WOLFRAM W. F. et al. [633], Seite 83–94. (Zitiert auf den Seiten 16 und 40)
- [160] Politzer, Thomas: *Vision Is Our Dominant Sense*. BrainLine, 2009. (Zitiert auf Seite 17)
- [161] Wade, Nicholas: *Visual Perception*. Psychology Press, 2001. (Zitiert auf Seite 17)
- [162] Herrmann, Christian, Juergen Metzler und Dieter Willersinn: *Semi-automatic people counting in aerial images of large crowds*. In: *SourceSociety of Photo-Optical Instrumentation Engineers*, Band 8542, Seite 85420Q–85420Q–8, 2012. (Zitiert auf den Seiten 17 und 23)
- [163] Greenshields, B. D.: *The Photographic Method of studying Traffic Behaviour*. In: *Proceedings of the 13th Annual Meeting of the Highway Research Board*, 1933. (Zitiert auf Seite 17)
- [164] Board, Transportation Research: *75 Years of the Fundamental Diagram for Traffic Flow Theory: Greenshields Symposium*. Nummer E-C149 in *Transportation Research E-Circular*. Transportation Research Board, 2011. (Zitiert auf Seite 17)
- [165] Stauffer, C. und W.E.L. Grimson: *Adaptive background mixture models for real-time tracking*. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Seite 246–252, 1999. (Zitiert auf Seite 18)
- [166] Wang, H. und D. Suter: *A re-evaluation of mixture of gaussian background modeling*. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Band 2, Seite 1017–1020, 2005. (Zitiert auf Seite 18)
- [167] Intille, Stephen S., James W. Davis und Aaron F. Bobick: *Real-time closed-world tracking*. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Seite 697–703, 1997. (Zitiert auf den Seiten 18 und 23)
- [168] Bertozzi, M., E. Binelli, A. Broggi und M. Del Rose: *Stereo vision-based approaches for pedestrian detection*. In: *IEEE International Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum*, 2005. (Zitiert auf Seite 18)
- [169] McKenna, S. J., S. Jabri, Z. Duric und H. Wechsler: *Tracking interacting people*. In: *IEEE International Conference on Automatic Face and Gesture Recognition*, Seite 348–454, 2000. (Zitiert auf Seite 18)

- [170] Zhang, X. und G. Sexton: *A new method for pedestrian counting*. In: *International Conference on Image Processing and its Applications*, Seite 208–212, 1995. (Zitiert auf den Seiten 18 und 23)
- [171] Beardsley, P. und E. Bourrat: *Wheelchair detection using stereo vision*. Technical Report, MERL, 2002. (Zitiert auf Seite 18)
- [172] Maurin, B., O. Masoud und N. P. Papanikolopoulos: *Camera surveillance of crowded traffic scenes*. In: *ITS America Annual Meeting*, Seite 28–55. Intelligent Transportation Society of America, 2002. (Zitiert auf Seite 18)
- [173] Horn, B. K. P. und B. G. Rhunck: *Determining optical flow: a retrospective*. *Artificial Intelligence*, 59:81–87, 1993. (Zitiert auf Seite 18)
- [174] Ali, A. T. und E. L. Dagless: *Vehicle and pedestrian detection and tracking*. In: *IEE Colloquium on Image Analysis for Transport Applications*, Seite 48–54, 1990. (Zitiert auf Seite 18)
- [175] Masoud, O. und N. P. Papanikolopoulos: *A novel method for tracking and counting pedestrians in real-time using a single camera*. *IEEE Transactions on Vehicular Technology*, 50:1267–1278, 2001. (Zitiert auf Seite 19)
- [176] Masoud, Osama und Nikolaos P. Papanikolopoulos: *Robust pedestrian tracking using a model-based approach*. In: *Intelligent Transportation System, 1997. ITSC'97., IEEE Conference on*, Seite 338–343, 1997. (Zitiert auf Seite 19)
- [177] Snidaro, L., C. Micheloni und C. Chiavedale: *Video security for ambient intelligence*. *Systems, Man and Cybernetics, Part A: Systems and Humans*, *IEEE Transactions on*, 35(1):133–144, 2005. (Zitiert auf den Seiten 19 und 23)
- [178] Lefloch, Damien: *Real-Time People Counting system using Video Camera*. Diplomarbeit, Gjøvik University College, 2007. (Zitiert auf den Seiten 19 und 23)
- [179] Gavrilu, D.M. und S. Munder: *Multi-cue Pedestrian Detection and Tracking from a Moving Vehicle*. *International Journal of Computer Vision*, 73(1):41–59, 2007. (Zitiert auf den Seiten 19, 20 und 40)
- [180] Felzenszwalb, Pedro, David McAllester und Deva Ramanan: *A discriminatively trained, multiscale, deformable part model*. In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, Seite 1–8, 2008. (Zitiert auf den Seiten 20 und 21)
- [181] Delamarre, Q. und O. Faugeras: *3d articulated models and multi-view tracking with silhouettes*. In: *International Conference on Computer Vision*, Band 2, Seite 716–721, 1999. (Zitiert auf den Seiten 20, 21 und 33)
- [182] Fotiadis, Efstathios P., Mario Garzón und Antonio Barrientos: *Human Detection from a Mobile Robot Using Fusion of Laser and Vision Information*. *Sensors*, 13(9):11603–11635, 2013. (Zitiert auf den Seiten 20 und 40)
- [183] Gavrilu, D. M. und V. Philomin: *Real-time object detection for smart vehicles*. In: *IEEE International Conference on Computer Vision*, Band 1, Seite 87–93, 1999. (Zitiert auf Seite 19)
- [184] Gavrilu, D. M. und J. Giebel: *Shape-based pedestrian detection and tracking*. In: *IEEE Intelligent Vehicles Symposium*, Band 1, Seite 8–14, 2002. (Zitiert auf Seite 19)

- [185] Baumberg, A.: *Learning Deformable Models for Tracking Human Motion*. Dissertation, University of Leeds, 1995. (Zitiert auf Seite 19)
- [186] Philomin, V., R. Duraiswami und L.S. Davis: *Pedestrian tracking from a moving vehicle*. In: *IEEE Intelligent Vehicles Symposium*, Seite 350–355, 2000. (Zitiert auf Seite 19)
- [187] Siebel, Nils T. und Steve J. Maybank: *Fusion of Multiple Tracking Algorithms for Robust People Tracking*. In: *Proceedings of European Conference on Computer Vision (ECCV) 02*, Seite 373–387, 2002. (Zitiert auf Seite 19)
- [188] Remagnino, P., A. Baumberg, T. Grove, D. Hogg, T. Tan, A. Worrall und K. Baker: *An Integrated Traffic and Pedestrian Model-Based Vision System*. In: *British Machine Vision Conference*, Band 2, Seite 380–389, 1997. (Zitiert auf Seite 19)
- [189] Sidla, O., Y. Lypetsky, N. Brandle und S. Seer: *Pedestrian detection and tracking for counting applications in crowded situations*. In: *IEEE International Conference on Video and Signal Based Surveillance*, Seite 70–75, 2006. (Zitiert auf den Seiten 19 und 23)
- [190] Vapnik, Vladimir: *The nature of statistical learning theory*. Springer, New York, 2000. (Zitiert auf Seite 20)
- [191] Piao, Songlin und Karsten Berns: *Vision Based Person Detection for Safe Navigation of Commercial Vehicle*. In: *Proceedings of the 13th International Conference on Intelligent Autonomous Systems (IAS-13)*, 2014. (Zitiert auf Seite 20)
- [192] Kong, Xiao-fang, Xiu-qin Wang, Guohua Gu, Qian Chen und Wei-xian Qian: *Human recognition based on head-shoulder contour extraction and BP neural network*. In: *International Symposium on Optoelectronic Technology and Application*, Band 9301, Seite 93010Z–93010Z–9, 2014. (Zitiert auf Seite 20)
- [193] Freund, Yoav und Robert E Schapire: *A decision-theoretic generalization of on-line learning and an application to boosting*. In: *Computational learning theory*, Seite 23–37, 1995. (Zitiert auf Seite 20)
- [194] Breiman, Leo: *Random forests*. *Machine learning*, 45(1):5–32, 2001. (Zitiert auf Seite 20)
- [195] Zhao, Liang und C. Thorpe: *Stereo- and neural network-based pedestrian detection*. In: *Intelligent Transportation Systems, 1999. Proceedings. 1999 IEEE/IEEJ/JSAI International Conference on*, Seite 298–303, 1999. (Zitiert auf den Seiten 20, 21 und 27)
- [196] Felzenszwalb, Pedro F., Ross B. Girshick und David McAllester: *Cascade object detection with deformable part models*. In: *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*, Seite 2241–2248, 2010. (Zitiert auf Seite 20)
- [197] Vedaldi, Andrea, Varun Gulshan, Manik Varma und Andrew Zisserman: *Multiple kernels for object detection*. In: *Computer Vision, 2009 IEEE 12th International Conference on*, Seite 606–613, 2009. (Zitiert auf Seite 20)
- [198] Viola, Paul und Michael J Jones: *Robust real-time face detection*. *International journal of computer vision*, 57(2):137–154, 2004. (Zitiert auf Seite 20)

- [199] Li, B., Y. Chen und F. Wang: *Pedestrian Detection based on Clustered Poselet Models and Hierarchical And-Or Grammar*. Vehicular Technology, IEEE Transactions on, PP(99):1–1, 2014. (Zitiert auf Seite 20)
- [200] Dalal, Navneet und Bill Triggs: *Histograms of Oriented Gradients for Human Detection*. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*, Seite 886–893, 2005. (Zitiert auf Seite 21)
- [201] Leibe, Bastian: *Figure-Ground Segmentation - Object-Based*. In: MOESLUND, THOMAS B. et al. [108], Seite 53–70. (Zitiert auf Seite 21)
- [202] Lowe, David G.: *Distinctive image features from scale-invariant keypoints*. International journal of computer vision, 60(2):91–110, 2004. (Zitiert auf Seite 21)
- [203] Luo, Ping, Yonglong Tian, Xiaogang Wang und Xiaoou Tang: *Switchable Deep Network for Pedestrian Detection*. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seite 899–906, 2014. (Zitiert auf Seite 21)
- [204] Leibe, Bastian, Edgar Seemann und Bernt Schiele: *Pedestrian detection in crowded scenes*. In: *Computer Vision and Pattern Recognition 2005*, Seite 878–885, 2005. (Zitiert auf den Seiten 21, 22 und 23)
- [205] Leibe, Bastian, Aleš Leonardis und Bernt Schiele: *Robust object detection with interleaved categorization and segmentation*. International journal of computer vision, 77(1-3):259–289, 2008. (Zitiert auf Seite 21)
- [206] García-Martín Álvaro, Alexander Hauptmann und Jose M. Martinez: *People Detection Based on Appearance and Motion Models*. In: *8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS) 2011*, Seite 5, 2011. (Zitiert auf Seite 21)
- [207] Chen, M.-Y. und A. Hauptmann: *Mosift: Recognizing human actions in surveillance videos*. Technical Report CMU-CS-09-161, Carnegie Mellon University, 2009. (Zitiert auf Seite 21)
- [208] Gavrilu, D. M. und L.S. Davis: *3d model-based tracking of human upper body movement: a multi-view approach*. In: *IEEE International Symposium on Computer Vision*, Seite 253–258, 1995. (Zitiert auf den Seiten 21 und 33)
- [209] Zhao, T. und R. Nevatia: *Bayesian human segmentation in crowded situations*. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2003*, Band II der Reihe *Proceedings - IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Seite 459–466, 2003. (Zitiert auf den Seiten 21, 23 und 27)
- [210] Idrees, Haroon, Imran Saleemi, Cody Seibert und Mubarak Shah: *Multi-source Multi-scale Counting in Extremely Dense Crowd Images*. In: *Conference on Computer Vision and Pattern Recognition*, Seite 2547–2554, 2013. (Zitiert auf den Seiten 21 und 23)
- [211] Horbert, Esther, Konstantinos Rematas und Bastian Leibe: *Level-Set Person Segmentation and Tracking with Multi-Region Appearance Models and Top-Down Shape Information*. In: *Proceeding of IEEE International Conference on Computer Vision (ICCV) [630]*, Seite 1871–1878. (Zitiert auf Seite 21)

- [212] Pishchulin, Leonid, Arjun Jain, Christian Wojek, Thorsten Thormählen und Bernt Schiele: *In Good Shape: Robust People Detection based on Appearance and Shape*. In: *British Machine Vision Conference (BMVC)*, 2011. (Zitiert auf Seite 21)
- [213] Micilotta, A., E. Ong und R. Bowden: *Detection and tracking of humans by probabilistic body part assembly*. In: *British Machine Vision Conference*, Band 1, Seite 419–428, 2005. (Zitiert auf Seite 21)
- [214] Mikolajczyk, K., C. Schmid und A. Zisserman: *Human detection based on a probabilistic assembly of robust part detectors*. In: *European Conference on Computer Vision*, Band 1, Seite 69–81, 2004. (Zitiert auf Seite 21)
- [215] Russell, James: *Detecting Humans in Video Footage using Multiple Classifiers*. Dissertation, School of Computer Science and Software Engineering, The University of Western Australia, 2004. (Zitiert auf Seite 21)
- [216] Viola, Paul, Michael J. Jones und Daniel Snow: *Detecting Pedestrians Using Patterns of Motion and Appearance*. In: *International Conference on Computer Vision*, Seite 734–741, 2003. (Zitiert auf Seite 21)
- [217] Leibe, Bastian, Konrad Schindler und Luc Van Gool: *Coupled detection and trajectory estimation for multi-object tracking*. In: *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, Seite 1–8, 2007. (Zitiert auf Seite 21)
- [218] Liu, Yazhou, Xilin Chen, Hongxun Yao, Xinyi Cui, Chaoran Liu und Wen Gao: *Contour-motion feature (CMF): A space-time approach for robust pedestrian detection*. *Pattern Recognition Letters*, 30(2):148–156, 2009. (Zitiert auf Seite 21)
- [219] Andriluka, Mykhaylo, Stefan Roth und Bernt Schiele: *People-Tracking-by-Detection and People-Detection-by-Tracking*. In: *IEEE Conference on Computer Vision and Pattern Recognition 2008*, 2008. (Zitiert auf Seite 21)
- [220] Cutler, Ross und Larry S. Davis: *Robust real-time periodic motion detection, analysis, and applications*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):781–796, 2000. (Zitiert auf Seite 21)
- [221] Pai, Chia-Jung, Hsiao-Rong Tyan, Yu-Ming Liang, Hong-Yuan Mark Liao und Sei-Wang Chen: *Pedestrian detection and tracking at crossroads*. *Pattern Recognition*, 37(5):1025–1034, 2004. (Zitiert auf Seite 21)
- [222] Niyogi, Sourabh A. und Edward H. Adelson: *Analyzing and recognizing walking figures in XYT*. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Seite 469–474, 1994. (Zitiert auf Seite 21)
- [223] Heisele, B und C Wohler: *Motion-based recognition of pedestrians*. In: Jain, AK und Venkatesh, S und Lovell, BC (Herausgeber): *14th International Conference on Pattern Recognition, Vols 1 and 2*, International Conference on Pattern Recognition, Seite 1325–1330, 1998. (Zitiert auf Seite 21)
- [224] Dollár, Piotr, Christian Wojek, Bernt Schiele, Pietro Perona und Tu Darmstadt: *Pedestrian detection: A benchmark*. In: *Conference on Computer Vision and Pattern Recognition*, 2009. (Zitiert auf Seite 22)

- [225] Schwartz, William Robson, Aniruddha Kembhavi, David Harwood und Larry S. Davis: *Human detection using partial least squares analysis*. In: *Computer Vision, 2009 IEEE 12th International Conference on*, Seite 24–31, 2009. (Zitiert auf Seite 22)
- [226] Tu, Peter H., Thomas B. Sebastian, Gianfranco Doretto, Nils O. Krahnstoeve, Jens Rittscher und Ting Yu: *Unified crowd segmentation*. In: *European Conference on Computer Vision*, Seite 691–704, 2008. (Zitiert auf den Seiten 22 und 23)
- [227] Cheriadat, Anil M., Budhendra L. Bhaduri und Richard J. Radke: *Detecting multiple moving objects in crowded environments with coherent motion regions*. In: *Computer Vision and Pattern Recognition Workshops 2008*, Band 0, Seite 1–8, 2008. (Zitiert auf den Seiten 22, 23 und 223)
- [228] Le Bon, Gustave: *The Crowd: A Study of the Popular Mind (Psychologie des Foules)*. Sparkling Books, 1895. (Zitiert auf Seite 23)
- [229] Dehghan, Afshin, Haroon Idrees, AmirRoshan Zamir und Mubarak Shah: *Automatic Detection and Tracking of Pedestrians in Videos with Various Crowd Densities*. In: WEIDMANN, ULRICH et al. [631], Seite 3–19. (Zitiert auf Seite 23)
- [230] Still, G. Keith: *Introduction to Crowd Science*. CRC Press, 2014. (Zitiert auf Seite 23)
- [231] Butenuth, Matthias, Florian Burkert, Florian Schmidt, Stefan Hinz, Dirk Hartmann, Angelika Kneidl, André Borrmann und Beril Sirmacek: *Integrating pedestrian simulation, tracking and event detection for crowd analysis*. In: *Proceeding of IEEE International Conference on Computer Vision (ICCV), Workshops*, Seite 150–157, 2011. (Zitiert auf Seite 23)
- [232] Rodriguez, Mikel, Josef Sivic, Ivan Laptev und Jean-Yves Audibert: *Density-aware person detection and tracking in crowds*. In: *Proceedings of the International Conference on Computer Vision (ICCV) [630]*. (Zitiert auf Seite 23)
- [233] Rodriguez, Mikel, Josef Sivic, Ivan Laptev und Jean-Yves Audibert: *Data-driven Crowd Analysis in Videos*. In: *Proceedings of the IEEE International Conference on Computer Vision (2011) [630]*, Seite 1235–1242. (Zitiert auf Seite 23)
- [234] Zhaoa, Tao, Ram Nevatia und Bo Wu: *Segmentation and Tracking of Multiple Humans in Crowded Environments*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(7):1198–1211, 2008. (Zitiert auf Seite 23)
- [235] Moore, Brian E., Saad Ali, Ramin Mehran und Mubarak Shah: *Visual Crowd Surveillance Through a Hydrodynamics Lens*. *Communications of the ACM*, 54(12):64–73, 2011. (Zitiert auf Seite 23)
- [236] Ali, Saad und Mubarak Shah: *Floor Fields for Tracking in High Density Crowd Scenes*. In: *10th European Conference on Computer Vision (ECCV)*, 2008. (Zitiert auf Seite 23)
- [237] Lypetsky, Yuriy: *Robust pedestrian detection and tracking in crowded scenes*. In: *Proc. SPIE 6764, Intelligent Robots and Computer Vision XXV: Algorithms, Techniques, and Active Vision*, Band 6764, 2007. (Zitiert auf Seite 23)
- [238] Chan, A.B., Z.-S.J. Liang und N. Vasconcelos: *Privacy preserving crowd monitoring: Counting people without people models or tracking*. In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, Seite 1–7, 2008. (Zitiert auf Seite 23)

- [239] Solmaz, Berkan, Brian E. Moore und Mubarak Shah: *Identifying Behaviors in Crowd Scenes Using Stability Analysis for Dynamical Systems*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 34(10):2064–2070, 2012. (Zitiert auf Seite 23)
- [240] Szczodrak, Maciej und Andrzej Czyzewski: *Video analytics-based algorithm for monitoring egress from buildings*. Multimedia Tools and Applications, Seite 1–11, 2014. (Zitiert auf Seite 23)
- [241] Dridi, M. H.: *Tracking Individual Targets in High Density Crowd Scenes Analysis of a Video Recording in Hajj 2009*. arXiv e-prints, 2014. <http://arxiv.org/abs/1407.2044> (10.07.2014). (Zitiert auf Seite 23)
- [242] Helbing, D., A. Johansson, H. Z. Al-Abideen und S. Al-Bosta: *Evaluations of video data of crowds around the Jamarat Bridge and safety implications for mass events*. In: *European Conference on Complex Systems*, 2007. (Zitiert auf Seite 23)
- [243] Tu, Peter H. und Jens Rittscher: *Crowd Segmentation Through Emergent Labeling*. In: Comaniciu, Dorin, Rudolf Mester, Kenichi Kanatani und David Suter (Hrsg.): *Statistical Methods in Video Processing*, Band 3247 der Reihe *Lecture Notes in Computer Science*, Seite 187–198. Springer Berlin Heidelberg, 2004. (Zitiert auf Seite 23)
- [244] Albiol, A., I. Mora und V. Naranjo: *Real-time high density people counter using morphological tools*. Intelligent Transportation Systems, IEEE Transactions on, 2(4):204–218, 2001. (Zitiert auf Seite 23)
- [245] Crabtree, M. R.: *Smart pedestrian counter system*. In: *International Conference on Road Transport Information and Control*, Seite 100–104, 2002. (Zitiert auf Seite 23)
- [246] Holzbach, Gerrit: *Personendetektion aus der Vogelperspektive mit einem Fischaugenobjektiv*. Technical Report, Fakultät für Informatik, Institut für Anthropomatik, 2013. (Zitiert auf Seite 23)
- [247] Busin, Laurent, Nicolas Vandenbroucke, Ludovic Macaire et al.: *Color spaces and image segmentation*. Advances in imaging and electron physics, 151:66, 2008. (Zitiert auf Seite 25)
- [248] Fukunaga, K. und C. D. Hosteller: *The estimation of the gradient of a density function, with application in pattern recognition*. In: *IEEE Transactions on Information Theory*, Band 21, Seite 32–40, 1975. (Zitiert auf Seite 25)
- [249] Comaniciu, D. und P. Meer: *Mean shift mode seeking and clustering*. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Band 17, Seite 790–799, 2002. (Zitiert auf Seite 25)
- [250] Christoudias, Christopher M.: *Synergism in Low Level Vision*. In: *Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02) Volume 4 - Volume 4*, ICPR '02, Seite 40150–, 2002. (Zitiert auf Seite 25)
- [251] Bakstein, Hynek, Michal Havlena, Petr Pohl und Tomáš Pajdla: *Omnidirectional sensors and their calibration for the Dirac project*. Technical Report, Research Report CTU–CMP–2006–13, CMP Prague, 2006. (Zitiert auf Seite 26)
- [252] Fiala, Mark: *ARTag, a fiducial marker system using digital techniques*. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Band 2, Seite 590–596, 2005. (Zitiert auf Seite 26)

- [253] Daniel, W. und S. Dieter: *Artoolkit for pose tracking on mobile devices*. In: *Proceedings of 12th Computer Vision Winter Workshop*, 2007. (Zitiert auf Seite 26)
- [254] Technology, Graz University of: *ARToolKitPlus*. http://studierstube.icg.tugraz.at/handheld_ar/artoolkitplus.php (12.05.2014). (Zitiert auf Seite 26)
- [255] Scaramuzza, D., A. Martinelli und R. Siegwart: *A Flexible Technique for Accurate Omnidirectional Camera Calibration and Structure from Motion*. In: *Computer Vision Systems, 2006 ICVS '06. IEEE International Conference on*, Seite 45–45, 2006. (Zitiert auf Seite 26)
- [256] Scaramuzza, D., A. Martinelli und R. Siegwart: *A Toolbox for Easily Calibrating Omnidirectional Cameras*. In: *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, Seite 5695–5701, 2006. (Zitiert auf Seite 26)
- [257] Ruffi, M., D. Scaramuzza und R. Siegwart: *Automatic detection of checkerboards on blurred and distorted images*. In: *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, Seite 3121–3126, 2008. (Zitiert auf Seite 26)
- [258] Foxlin, Eric und Leonid Naimark: *VIS-Tracker: A Wearable Vision-Inertial Self-Tracker*. In: *Proceedings of the IEEE Virtual Reality 2003, VR '03*, Seite 199–, 2003. (Zitiert auf Seite 26)
- [259] Onishi, Masaki und Ikushi Yoda: *Dynamic Trajectory Extraction from Stereo Vision Using Fuzzy Clustering*. *IEEJ Transactions on Electronics, Information and Systems*, 128(9):1438–1446, 2008. (Zitiert auf Seite 26)
- [260] Zhao, Liang: *Dressed human modeling, detection, and parts localization*. Dissertation, Carnegie Mellon University, 2001. (Zitiert auf Seite 26)
- [261] Kelly, Philip, Noel E. O'Connor und Alan F. Smeaton: *A Framework for Evaluating Stereo-Based Pedestrian Detection Techniques*. *IEEE Transactions On Circuits And Systems For Video Technology*, 18:1163–1167, 2008. (Zitiert auf Seite 26)
- [262] Ding, Jianwei, Yongzhen Huang, Kaiqi Huang und Tieniu Tan: *Robust Object Tracking via Online Learning of Adaptive Appearance Manifold*. In: *Proceeding of IEEE International Conference on Computer Vision (ICCV) [630]*, Seite 1863–1869. (Zitiert auf Seite 26)
- [263] Scovanner, Paul und Marshall F. Tappen: *Learning Pedestrian Dynamics from the Real World*. In: *Proceeding of IEEE International Conference on Computer Vision (ICCV)*, 2009. (Zitiert auf Seite 26)
- [264] Rojas, Raul: *Stereo vision with a single camera*. Technical Report, Freie Universität Berlin, 2006. (Zitiert auf Seite 27)
- [265] Beymer, D. und K. Konolige: *Real-Time Tracking of Multiple People Using Stereo*. In: *IEEE Frame Rate Workshop*, 1999. (Zitiert auf den Seiten 27 und 28)
- [266] Darrell, T, G Gordon, M Harville und J Woodfill: *Integrated person tracking using stereo, color, and pattern detection*. *International Journal of Computer Vision*, 37(2):175–185, 2000. (Zitiert auf den Seiten 27 und 40)
- [267] Gibson, David R. P., Bo Ling, Michael Zeifman, Shaoqiang Dong und Uma Venkataraman: *Multipedestrian tracking*. *Public Roads*, 69(5):8, 2006. (Zitiert auf Seite 27)

- [268] Luo, Ruijiang und Yan Guo: *Real-Time Stereo Tracking of Multiple Moving Heads*. IEEE International Conference on Computer Vision Workshop on Recognition, Analysis, & Tracking of Faces & Gestures in Real -Time Systems, 0:0055, 2001. (Zitiert auf Seite 27)
- [269] Zhao, L. und C. Thorpe: *Recursive context reasoning for human detection and parts identification*. In: *IEEE Workshop on Human Modeling, Analysis, and Synthesis*, 2000. (Zitiert auf Seite 27)
- [270] Haritaoglu, I., D. Harwood und L. S. Davis.: *W⁴s: A real time system for detecting and tracking people in 2.5d*. In: *European Conference on Computer Vision*, Seite 877–892, 1998. (Zitiert auf Seite 27)
- [271] Krumm, J, S Harris, B Meyers, B Brumitt, M Hale und S Shafer: *Multi-camera multi-person tracking for EasyLiving*. In: *Third IEEE International Workshop on Visual Surveillance*, Seite 3–10, 2000. (Zitiert auf den Seiten 27 und 33)
- [272] Elgammal, AM und LS Davis: *Probabilistic framework for segmenting people under occlusion*. In: *8th IEEE International Conference on Computer Vision*, Band II, Seite 145–152, 2001. (Zitiert auf Seite 27)
- [273] Rittscher, Jens, Peter H. Tu und Nils O. Krahnstoeber: *Simultaneous estimation of segmentation and shape*. In: Schmid, C. and Soatto, S. and Tomasi, C. (Herausgeber): *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2*, Seite 486–493, 2005. (Zitiert auf Seite 27)
- [274] Hu, WM, X Zhou, TN Tan, JG Lou und S Maybank: *Principal axis-based correspondence between multiple cameras for people tracking*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(4):663–671, 2006. (Zitiert auf den Seiten 27 und 33)
- [275] Salinas, Rafael Muñoz, Eugenio Aguirre und Miguel García-Silvente: *People detection and tracking using stereo vision and color*. Image Vision Computing, 25:995–1007, 2007. (Zitiert auf den Seiten 28, 29 und 30)
- [276] Walk, Stefan, Konrad Schindler und Bernt Schiele: *Disparity statistics for pedestrian detection: Combining appearance, motion and stereo*. In: *European Conference on Computer Vision (ECCV)*, Seite 182–195. Springer, 2010. (Zitiert auf Seite 28)
- [277] Nedeveschi, Sergiu, Silviu Bota und Corneliu Tomiuc: *Stereo-based pedestrian detection for collision-avoidance applications*. Intelligent Transportation Systems, IEEE Transactions on, 10(3):380–391, 2009. (Zitiert auf Seite 28)
- [278] Hou, Ya-Li und Grantham K. H. Pang: *Multi-cue-Based Crowd Segmentation in Stereo Vision*. In: Real, Pedro, Daniel Díaz-Pernil, Helena Molina-Abril, Ainhua Berciano und Walter G. Kropatsch (Hrsg.): *International Conference on Computer Analysis of Images and Patterns*, Band 6854 der Reihe *Lecture Notes in Computer Science*, Seite 93–101, 2011. (Zitiert auf Seite 28)
- [279] Bertozzi, M., A. Broggi, R. Chapuis, F. Chausse, A. Fascioli und A. Tibaldi: *Shape-based pedestrian detection and localization*. In: *IEEE International Conference on Intelligent Transportation Systems*, Seite 328–333, 2003. (Zitiert auf Seite 28)

- [280] Kim, K. und L. S. Davis: *Multi-camera tracking and segmentation of occluded people on ground plane using search-guided particle filtering*. In: *European Conference on Computer Vision*, Band 3, Seite 98–109, 2006. (Zitiert auf den Seiten 28 und 32)
- [281] Sudowe, Patrick und Bastian Leibe: *Efficient Use of Geometric Constraints for Sliding-Window Object Detection in Video*. In: *International Conference on Computer Vision Systems*, 2011. (Zitiert auf Seite 28)
- [282] Terada, K., D. Yoshida, S. Oe und J. Yamaguchi: *A counting method of the number of passing people using a stereo camera*. In: *IEEE Conference on Industrial Electronics*, Band 3, Seite 1318–1323, 1999. (Zitiert auf Seite 28)
- [283] Haritaoglu, I., D. Beymer und M. Flickner: *Ghost3d: Detecting body posture and parts using stereo*. In: *Workshop on Motion and Video Computing*, Seite 175–180, 2002. (Zitiert auf Seite 28)
- [284] Beymer, D.: *Person counting using stereo*. In: *Human Motion, 2000. Proceedings. Workshop on*, Seite 127–133, 2000. (Zitiert auf den Seiten 28 und 30)
- [285] Filipov, Rusi Antonov, Flávio Luis Cardeal Pádua und Marco Aurélio Bueno Carone: *Pylon grid: A fast method for human head detection in range images*. *Neurocomputing*, Special issue: Behaviours in video, 100(0):74 – 85, 2013. (Zitiert auf den Seiten 28, 30 und 31)
- [286] Oosterhout, Tim van, Sander Bakkes und Ben Kröse: *Head Detection in Stereo Data for People Counting and Segmentation*. In: *International Conference on Computer Vision Theory and Applications (VISAPP)*, Seite 620–625, 2011. (Zitiert auf den Seiten 28 und 30)
- [287] Zhang, Zhongchuan und Fernand Cohen: *Pedestrian Tracking based on 3D Head Point Detection*. In: *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Seite 382–385, 2013. (Zitiert auf Seite 28)
- [288] Zhang, Zhongchuan und Fernand Cohen: *3D pedestrian tracking based on overhead cameras*. In: *Seventh International Conference on Distributed Smart Cameras (ICDSC)*, Seite 1–6, 2013. (Zitiert auf Seite 28)
- [289] Serrano-Cuerda, Juan, José Carlos Castillo und Antonio Fernández-Caballero: *Indoor Overhead Video Camera for Efficient People Counting*. *Jurnal Teknologi*, 63(3):17–22, 2013. (Zitiert auf Seite 28)
- [290] Serrano-Cuerda, Juan, José Carlos Castillo, Marina V Sokolova und Antonio Fernández-Caballero: *Efficient people counting from indoor overhead video camera*. In: *Trends in Practical Applications of Agents and Multiagent Systems*, Seite 129–137. Springer, 2013. (Zitiert auf Seite 28)
- [291] Bajracharya, M., B. Moghaddam, A. Howard, S. Brennan und L. Matthies: *A Fast Stereo-based System for Detecting and Tracking Pedestrians from a Moving Vehicle*. *International Journal of Remote Sensing*, 28:11–12, 2009. (Zitiert auf Seite 28)
- [292] Bansal, M., S. H. Jung, B. Matei, J. Eledath und H. S. Sawhney: *A Real-time Pedestrian Detection System based on Structure and Appearance Classification*. In: *IEEE International Conference on Robotics and Automation (ICRA)*, 2010. (Zitiert auf Seite 28)

- [293] Vedaldi, A. und S. Soatto: *Quick Shift and Kernel Methods for Mode Seeking*. In: *European Conference on Computer Vision*, 2008. (Zitiert auf den Seiten 28 und 29)
- [294] Hayashi, K., M. Hashimoto, K. Sumi und K. Sasakawa: *Multiple-person tracker with a fixed slanting stereo camera*. In: *IEEE International Conference on Automatic Face and Gesture Recognition*, Seite 681–686, 2004. (Zitiert auf den Seiten 28 und 30)
- [295] Harville, Michael: *Stereo person tracking with adaptive plan-view templates of height and occupancy statistics*. *Image and Vision Computing*, 22(2):127–142, 2004. (Zitiert auf den Seiten 28, 30 und 100)
- [296] Kelly, Philip, Noel E. O’Connor und Alan F. Smeaton: *Robust pedestrian detection and tracking in crowded scenes*. *Image and Vision Computing*, 27(10):1445–1458, 2009. (Zitiert auf den Seiten 28, 30 und 119)
- [297] Kelly, P., N. O’Connor und A. Smeaton: *Pedestrian detection in uncontrolled environments using stereo and biometric information*. In: *ACM International Workshop on Video Surveillance and Sensor Networks*, Seite 161–170, 2006. (Zitiert auf Seite 28)
- [298] Mitzel, Dennis und Bastian Leibe: *Close-Range Human Detection for Head-Mounted Cameras*. In: *British Machine Vision Conference*, 2012. (Zitiert auf den Seiten 29 und 30)
- [299] Harville, Michael: *Stereo Person Tracking with Adaptive Plan-View Statistical Templates*. Technical Report HPL-2002-122, Hewlett-Packard, 2002. (Zitiert auf Seite 30)
- [300] Harville, M. und L. Dalong: *Fast, integrated person tracking and activity recognition with plan-view templates from a single stereo camera*. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Seite 398–405, 2002. (Zitiert auf Seite 30)
- [301] Salinas, Rafael Muñoz, Eugenio Aguirre, Miguel García-Silvente und Antonio Gonzalez: *People Detection and Tracking Through Stereo Vision for Human-Robot Interaction*. In: Gelbukh, Alexander, Álvaro de Albornoz und Hugo Terashima-Marín (Hrsg.): *MICAI 2005: Advances in Artificial Intelligence*, Band 3789 der Reihe *Lecture Notes in Computer Science*, Seite 337–346. Springer Berlin Heidelberg, 2005. (Zitiert auf Seite 30)
- [302] Muñoz-Salinas, Rafael, R Medina-Carnicer, Francisco José Madrid-Cuevas und A Carmona-Poyato: *People detection and tracking with multiple stereo cameras using particle filters*. *Journal of Visual Communication and Image Representation*, 20(5):339–350, 2009. (Zitiert auf den Seiten 30 und 33)
- [303] Park, Sangho und Jake K. Aggarwal: *Head Segmentation and Head Orientation in 3D Space for Pose Estimation of Multiple People*. In: *IEEE Southwest Symposium on Image Analysis and Interpretation 2000*, Seite 192–196, 2000. (Zitiert auf Seite 30)
- [304] Mikić, Ivana, Mohan Trivedi, Edward Hunter und Pamela Cosman: *Human body model acquisition and motion capture using voxel data*. In: *Articulated Motion and Deformable Objects*, Seite 104–118. Springer, 2002. (Zitiert auf Seite 30)
- [305] Karpagavalli, P., A. Vinoth Nelson und A. V. Ramprasad: *Human Tracking and Counting In Range Images by Using Mean Shift and Pylon Grid Algorithm*. *International Journal of Innovative Research in Science, Engineering and Technology*, 3:1231–1237, 2014. (Zitiert auf Seite 31)

- [306] Hirschmüller, Heiko: *Stereo Processing by Semiglobal Matching and Mutual Information*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30:328–341, 2008. (Zitiert auf Seite 31)
- [307] Won, Kwang Hee, Sisay Gurmu und Soon Ki Jung: *Pedestrian detection using labeled depth data*. In: *Frontiers of Computer Vision, (FCV), 2013 19th Korea-Japan Joint Workshop on*, Seite 117–120, 2013. (Zitiert auf Seite 31)
- [308] Research, Point Grey: *Digiclops three-camera Stereo Vision system*. www.ptgrey.com/products/digiclops/Digiclops.pdf (15.06.2014). (Zitiert auf Seite 32)
- [309] Baltieri, Davide, Roberto Vezzani, Rita Cucchiara, Akos Utasi, Csaba Benedek und Tamas Sziranyi: *Multi-View People Surveillance Using 3D Information*. In: *Proceeding of IEEE International Conference on Computer Vision (ICCV) [630]*, Seite 1817–1824. (Zitiert auf Seite 32)
- [310] Sternig, Sabine, Thomas Mauthner, Arnold Irschara, Peter M. Roth und Horst Bischof: *Multi-camera Multi-object Tracking by Robust Hough-based Homography Projections*. In: *Proceeding of IEEE International Conference on Computer Vision (ICCV) [630]*, Seite 1689–1696. (Zitiert auf Seite 32)
- [311] Gall, Juergen und Victor Lempitsky: *Class-specific hough forests for object detection*. In: *Decision Forests for Computer Vision and Medical Image Analysis*, Seite 143–157. Springer, 2013. (Zitiert auf Seite 32)
- [312] Okada, R.: *Discriminative generalized hough transform for object detection*. In: *Computer Vision, 2009 IEEE 12th International Conference on*, Seite 2000–2005, 2009. (Zitiert auf Seite 32)
- [313] Lien, Kuo-Chin und Chung-Lin Huang: *Multiview-Based Cooperative Tracking of Multiple Human Objects*. EURASIP Journal on Image and Video Processing, 2008(1):253039, 2008. (Zitiert auf Seite 32)
- [314] Eshel, Ran und Yael Moses: *Homography based multiple camera detection and tracking of people in a dense crowd*. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 0:1–8, 2008. (Zitiert auf Seite 33)
- [315] Sivapalan, Sabesan, Daniel C.Y. Chen, Simon Denman, Sridha Sridharan und Clinton Fookes: *3D Ellipsoid Fitting for Multi-view Gait Recognition*. In: *8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS) 2011*, Seite 6, 2011. (Zitiert auf Seite 33)
- [316] Bouguet, J.-Y.: *Camera calibration toolbox for MATLAB*, 2008. http://www.vision.caltech.edu/bouguetj/calib_doc/ (12.06.2014). (Zitiert auf Seite 33)
- [317] Berclaz, Jerome, Francois Fleuret und Pascal Fua: *Robust people tracking with global trajectory optimization*. In: *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, Band 1, Seite 744–750, 2006. (Zitiert auf Seite 33)
- [318] Ge, Weina und Robert T. Collins: *Crowd detection with a multiview sampler*. In: *11th European conference on Computer vision: Part V, ECCV'10*, Seite 324–337, 2010. (Zitiert auf Seite 33)

- [319] Mörzinger, Roland, Marcus Thaler, Severin Stalder, Helmut Grabner und Luc Van Gool: *Improved Person Detection in Industrial Environments using Multiple Self-Calibrated Cameras*. In: *2011 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, Seite 6, 2011. (Zitiert auf Seite 33)
- [320] Utasi, A. und C. Benedek: *A Bayesian Approach on People Localization in Multicamera Systems*. *Circuits and Systems for Video Technology*, IEEE Transactions on, 23(1):105–115, 2013. (Zitiert auf Seite 33)
- [321] Kiss Ákos und Tamás Szirányi: *Localizing people in multi-view environment using height map reconstruction in real-time*. *Pattern Recognition Letters*, 34(16):2135 – 2143, 2013. (Zitiert auf Seite 33)
- [322] Zhang, Yongjun, Yong Zhao, Guoliang Li und Ruzhong Cheng: *Grey Self-similarity Feature for Night-time Pedestrian Detection*. *Journal of Computational Information Systems*, 10(7):2967–2974, 2014. (Zitiert auf Seite 34)
- [323] Zhang, X. L., W. G. Weng, H. Y. Yuan und J. G. Chen: *Empirical study of a unidirectional dense crowd during a real mass event*. *Physica A: Statistical Mechanics and its Applications*, 392(12):2781 – 2791, 2013. (Zitiert auf Seite 34)
- [324] Hadi, H. Saipol, M. Rosbi und U. U. Sheikh: *A Review of Infrared Spectrum in Human Detection for Surveillance Systems*. *International Journal Of Interactive Digital Media*, 1(3):13–20, 2013. (Zitiert auf Seite 34)
- [325] *Pedigree project*. <http://www.math.univ-toulouse.fr/pedigree/>. (Zitiert auf den Seiten 34, 56 und 181)
- [326] ZDF: *Schlau im Stau*, 2009. Abenteuer Wissen (12.08.2009). (Zitiert auf Seite 34)
- [327] Ltd., Vicon Motion Systems: *Vicon*. <http://www.vicon.com/> (25.06.2014). (Zitiert auf Seite 34)
- [328] Saito, Hiroaki, Takeshi Hagihara, Kenichi Hatanaka und Takanori Sawai: *Development of Pedestrian Detection System Using Far-Infrared Ray Camera*. *Sei Technical Review*, 66:112–117, 2008. (Zitiert auf Seite 35)
- [329] Nanda, Harsh und Larry Davis: *Probabilistic template based pedestrian detection in infrared videos*. In: *IEEE Intelligent Vehicle Symposium*, Band 1, Seite 15–20, 2002. (Zitiert auf Seite 35)
- [330] Kuutti, Jussi, Kim H Blomqvist, Raimo E Sepponen, Jaeyoung Kwak und Isakki Kosonen: *Performance of commercial over-head camera sensors in recognizing patterns of two and three persons: A case study*. In: *Applied Electrical Engineering and Computing Technologies (AEECT), 2013 IEEE Jordan Conference on*, Seite 1–6, 2013. (Zitiert auf Seite 35)
- [331] Wang, Jiang-Tao, De-Bao Chen, Hai-Yan Chen und Jing-Yu Yang: *On Pedestrian Detection and Tracking in Infrared Videos*. *Pattern Recognition Letters*, 2011. (Zitiert auf Seite 35)
- [332] Kerridge, Jon, Stefan Keller, Tim Chamberlain und Neil Sumpter: *Collecting Pedestrian Trajectory Data In Real-time*. In: WALDAU, N. et al. [628]. (Zitiert auf Seite 35)

- [333] Armitage, A., D. Binnie, J. Kerridge und L. Lei: *Measuring pedestrian trajectories with low cost infrared detectors: Preliminary results*. In: *Pedestrian and Evacuation Dynamics 2003*, Seite 101–110, 2003. (Zitiert auf Seite 35)
- [334] Li, Guoliang, Yong Zhao, Daimeng Wei und Ruzhong Cheng: *Nighttime Pedestrian Detection Using Local Oriented Shape Context Descriptor*. In: *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering*, 2013. (Zitiert auf Seite 35)
- [335] Gade, Rikke und Thomas B. Moeslund: *Thermal Tracking of Sports Players*. *Sensors*, 14(8):13679–13691, 2014. (Zitiert auf Seite 35)
- [336] Lin, Chun-Fu, Sheng-Fuu Lin, Chi-Hung Hwang und Yu-Chieh Chen: *Real-time pedestrian detection system with novel thermal features at night*. In: *Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, 2014 IEEE International*, Seite 1329–1333, 2014. (Zitiert auf Seite 35)
- [337] Yu, Dahai, Junwei Han, Yibo Ye und Zhijun Fang: *A novel saliency detection framework for infrared thermal images*. In: *Orange Technologies (ICOT), 2014 IEEE International Conference on*, Seite 57–60, 2014. (Zitiert auf Seite 35)
- [338] Qi, Bin, Vijay John, Zheng Liu und Seiichi Mita: *Pedestrian detection from thermal images with a scattered difference of directional gradients feature descriptor*. In: *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*, Seite 2168–2173, 2014. (Zitiert auf Seite 35)
- [339] Kuutti, Jussi: *A Test Setup for Comparison of People Flow Sensors*. Diplomarbeit, Aalto University, 2012. (Zitiert auf Seite 36)
- [340] Wikipedia: *Entfernungsmessung*. <http://de.wikipedia.org/wiki/Entfernungsmessung> (12.07.2014). (Zitiert auf Seite 36)
- [341] Jähne, Bernd: *Digitale Bildverarbeitung*. Springer, 2005. (Zitiert auf den Seiten 36, 53 und 121)
- [342] Sagawa, Ryusuke, Hiroshi Kawasaki, Shota Kiyota und Ryo Furukawa: *Dense One-shot 3D Reconstruction by Detecting Continuous Regions with Parallel Line Projection*. In: *Proceeding of IEEE International Conference on Computer Vision (ICCV) [630]*, Seite 1911–1918. (Zitiert auf Seite 36)
- [343] Fossati, Andrea, Juergen Gall, Helmut Grabner, Xiaofeng Ren und Kurt Konolige (Hrsg.): *Consumer Depth Cameras for Computer Vision: Research Topics and Applications*. Springer Publishing Company, Incorporated, 2012. (Zitiert auf Seite 37)
- [344] Microsoft: *Kinect for Windows SDK*. <http://msdn.microsoft.com/en-us/library/hh855347.aspx> (12.06.2014). (Zitiert auf Seite 37)
- [345] Smisek, Jan, Michal Jancosek und Tomas Pajdla: *3D with Kinect*. In: *Proceeding of IEEE International Conference on Computer Vision (ICCV) [630]*, Seite 1154–1160. (Zitiert auf Seite 37)
- [346] Khoshelham, Kourosh und Sander Oude Elberink: *Accuracy and resolution of Kinect depth data for indoor mapping applications*. *Sensors*, 12:1437–1454, 2012. (Zitiert auf Seite 37)

- [347] Hsieh, Ching-Tang, Hui-Chun Wang, Yeh-Kuang Wu, Liung-Chun Chang und Tai-Ku Kuo: *A Kinect-based people-flow counting system*. In: *Intelligent Signal Processing and Communications Systems (ISPACS), 2012 International Symposium on*, Seite 146–150, 2012. (Zitiert auf Seite 37)
- [348] Wu, Shengyin, Shiqi Yu und Wensheng Chen: *An attempt to pedestrian detection in depth images*. In: *Intelligent Visual Surveillance (IVS), 2011 Third Chinese Conference on*, Seite 97–100, 2011. (Zitiert auf Seite 37)
- [349] Spinello, Luciano und Kai Oliver Arras: *People detection in rgb-d data*. In: *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, Seite 3838–3843, 2011. (Zitiert auf Seite 37)
- [350] Gerova, Silyana, Vasil Georgiev, Fabian Göcke, Lars Grotehenne, Dominik Hamann, Patrick Hühne, Ben Rabeler und Christian Schildwächter: *Funktionsweise, Eigenschaften und Anwendung der Microsoft Kinect in der Robotik*. Technical Report, Institut für Verteilte Systeme, Universität Magdeburg, 2013. (Zitiert auf Seite 37)
- [351] Charreyron, Samuel, Stewart Jackson und Luis F. Miranda-Moreno: *Toward a flexible system for pedestrian data collection with a Microsoft Kinect motion-sensing device*. Transportation Research Record: Journal of the Transportation Research Board, 2339(1):80–89, 2013. (Zitiert auf Seite 37)
- [352] Shotton, Jamie, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook und Richard Moore: *Real-time human pose recognition in parts from single depth images*. Communications of the ACM, 56(1):116–124, 2013. (Zitiert auf Seite 37)
- [353] Sung, Jaeyong, Colin Ponce, Bart Selman und Ashutosh Saxena: *Human activity detection from rgb-d images*. In: *AAAI workshop on Pattern, Activity and Intent Recognition (PAIR)*, Band 2, Seite 7, 2011. (Zitiert auf Seite 37)
- [354] Sturm, J., E. Bylow, F. Kahl und D. Cremers: *CopyMe3D: Scanning and Printing Persons in 3D*. In: *German Conference on Pattern Recognition (GCPR)*, 2013. (Zitiert auf Seite 37)
- [355] Kerl, C., J. Sturm und D. Cremers: *Dense Visual SLAM for RGB-D Cameras*. In: *Intelligent Robots and Systems (IROS), Proceedings of the IEEE/RSJ International Conference on*, 2013. (Zitiert auf Seite 37)
- [356] Newcombe, Richard A., Andrew J. Davison, Shahram Izadi, Pushmeet Kohli, Otmar Hilliges, Jamie Shotton, David Molyneaux, Steve Hodges, David Kim und Andrew Fitzgibbon: *Kinect-Fusion: Real-time dense surface mapping and tracking*. In: *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*, Seite 127–136, 2011. (Zitiert auf Seite 37)
- [357] Bylow, E., J. Sturm, C. Kerl, F. Kahl und D. Cremers: *Real-Time Camera Tracking and 3D Reconstruction Using Signed Distance Functions*. In: *Robotics: Science and Systems Conference (RSS)*, 2013. (Zitiert auf Seite 37)
- [358] Schindhelm, Corina Kim: *Evaluating SLAM Approaches for Microsoft Kinect*. In: *International Conference on Wireless and Mobile Communications (ICWMC)*, 2012. (Zitiert auf Seite 37)
- [359] Seer, Stefan, Norbert Brändle und Carlo Ratti: *Kinects and Human Kinetics: A New Approach for Studying Crowd Behavior*. arXiv e-prints, 2012. (Zitiert auf den Seiten 37 und 38)

- [360] Corbetta, Alessandro, Adrian Muntean, Federico Toschi und Kiamars Vafayi: *Parameter Estimation of Social Forces in Crowd Dynamics Models via a Probabilistic Method*. arXiv e-prints, 2014. (Zitiert auf den Seiten 37 und 38)
- [361] Corbetta, Alessandro, Luca Bruno, Adrian Muntean und Federico Toschi: *High statistics measurements of pedestrian dynamics*. arXiv e-prints, 2014. (Zitiert auf den Seiten 37 und 38)
- [362] Porzycki, Jakub, Marcin Mycek und Jarosław Wąs Robert Lubaś: *Dynamic Data-driven Simulation of Pedestrian Movement with Automatic Validation*. In: *Traffic and Granular Flow 2013*, 2015. (Zitiert auf den Seiten 37, 38 und 39)
- [363] Duda, Richard O., Peter E. Hart und David G. Stork: *Pattern classification*. John Wiley & Sons, 2012. (Zitiert auf Seite 37)
- [364] Bernardin, Keni, Alexander Elbs und Rainer Stiefelhagen: *Multiple object tracking performance metrics and evaluation in a smart room environment*. In: *Sixth IEEE International Workshop on Visual Surveillance, in conjunction with ECCV*, Band 90, Seite 91, 2006. (Zitiert auf den Seiten 38 und 120)
- [365] Caon, Maurizio, Yong Yue, Julien Tscherrig, Elena Mugellini und Omar Abou Khaled: *Context-aware 3d gesture interaction based on multiple kinects*. In: *AMBIENT 2011, The First International Conference on Ambient Computing, Applications, Services and Technologies*, Seite 7–12, 2011. (Zitiert auf Seite 38)
- [366] Ahmed, Naveed und Imran Nazir Junejo: *A System for 3D Video Acquisition and Spatio-Temporally Coherent 3D Animation Reconstruction using Multiple RGB-D Cameras*. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 6(2):113–128, 2013. (Zitiert auf Seite 38)
- [367] Or-El, Roy: *Kinect Depth Map*. Seminar in Advanced Topics in Computer Vision, Technion, 2013. (Zitiert auf Seite 38)
- [368] Kramer, Jeff, Matt Parker, Herrera C. Daniel, Florian Echtler und Nicolas Burrus: *Hacking the Kinect*. Springer, Apress, 2012. (Zitiert auf Seite 38)
- [369] Berger, Kai, Kai Ruhl, Yannic Schroeder, Christian Bruemmer, Alexander Scholz und Marcus A. Magnor: *Markerless Motion Capture using multiple Color-Depth Sensors*. In: *International Workshop on Vision, Modeling and Visualization*, Seite 317–324, 2011. (Zitiert auf Seite 38)
- [370] Tong, Jing, Jin Zhou, Ligang Liu, Zhigeng Pan und Hao Yan: *Scanning 3d full human bodies using kinects*. *Visualization and Computer Graphics, IEEE Transactions on*, 18(4):643–650, 2012. (Zitiert auf Seite 38)
- [371] Maimone, Andrew und Henry Fuchs: *Reducing interference between multiple structured light depth sensors using motion*. In: *Virtual Reality Short Papers and Posters (VRW), 2012 IEEE*, Seite 51–54, 2012. (Zitiert auf Seite 38)
- [372] Butler, D. Alex, Shahram Izadi, Otmar Hilliges, David Molyneaux, Steve Hodges und David Kim: *Shake'n'sense: reducing interference for overlapping structured light depth cameras*. In: *Proceedings of the Special Interest Group on Computer-Human Interaction (SIGCHI) Conference on Human Factors in Computing Systems*, Seite 1933–1936, 2012. (Zitiert auf Seite 38)

- [373] Maimone, Andrew und Henry Fuchs: *Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras*. In: *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, Seite 137–146, 2011. (Zitiert auf Seite 38)
- [374] Chen, Xiaofeng, Kristian Henrikson und Yinhai Wang: *Kinect-based Pedestrian Detection for Crowded Scenes*. In: *North American Travel Monitoring Exposition and Conference (NAT-MEC) 2014 - Improving Traffic Data Collection, Analysis, and Use*, to appear. (Zitiert auf Seite 38)
- [375] Su, Song-Zhi, Zhi-Hui Liu, Su-Ping Xu, Shao-Zi Li und Rongrong Ji: *Sparse auto-encoder based feature learning for human body detection in depth image*. *Signal Processing*, 2014. (Zitiert auf Seite 38)
- [376] Choi, Wongun, Caroline Pantofaru und Silvio Savarese: *Detecting and Tracking People using an RGB-D Camera via Multiple Detector Fusion*. In: *Proceeding of IEEE International Conference on Computer Vision (ICCV) [630]*, Seite 1076–1083. (Zitiert auf Seite 39)
- [377] Basso, Filippo, Matteo Munaro, Stefano Michieletto, Enrico Pagello und Emanuele Menegatti: *Fast and robust multi-people tracking from rgb-d data for a mobile robot*. In: *Intelligent Autonomous Systems 12*, Seite 265–276. Springer, 2013. (Zitiert auf Seite 39)
- [378] Luber, Matthias, Luciano Spinello und Kai Oliver Arras: *People tracking in rgb-d data with on-line boosted target models*. In: *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, Seite 3844–3849, 2011. (Zitiert auf Seite 39)
- [379] Xia, Lu, Chia-Chih Chen und J. K. Aggarwal: *Human detection using depth information by kinect*. In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, Seite 15–22, 2011. (Zitiert auf Seite 39)
- [380] Salas, Joaquín und Carlo Tomasi: *People Detection Using Color and Depth Images*. In: Martínez-Trinidad, José Francisco, Jesús Ariel Carrasco-Ochoa, Cherif Ben-Youssef Brants und Edwin Robert Hancock (Hrsg.): *Pattern Recognition*, Band 6718 der Reihe *Lecture Notes in Computer Science*, Seite 127–135. Springer Berlin Heidelberg, 2011. (Zitiert auf Seite 39)
- [381] Hussmann, Stephan, Thorsten Ringbeck und Bianca Hagebeuker: *A performance review of 3D TOF vision systems in comparison to stereo vision systems*. *Stereo Vision*, Seite 103–120, 2008. (Zitiert auf Seite 39)
- [382] Yarovoy, A. G., L. P. Ligthart, J. Matuzas und B. Levitas: *UWB radar for human being detection*. *Aerospace and Electronic Systems Magazine, IEEE*, 21(3):10–14, 2006. (Zitiert auf Seite 40)
- [383] Gürbüz, Sevgi Z., William L. Melvin und Douglas B. Williams: *Detection and identification of human targets in radar data*. In: *Defense and Security Symposium*, Seite 65670I–65670I, 2007. (Zitiert auf Seite 40)
- [384] Maaref, Nadia, Patrick Millot, Christian Pichot und Odile Picon: *A study of UWB FM-CW radar for the detection of human beings in motion inside a building*. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(5):1297–1300, 2009. (Zitiert auf Seite 40)

- [385] Heckenkamp, Christoph: *Das PMD-Prinzip*. Technical Report 4, Inspect, 2012. (Zitiert auf Seite 40)
- [386] Benedek, C.: *3D People Surveillance On Range Data Sequences Of A Rotating Lidar*. Pattern Recognition Letters, Special Issue on Depth Image Analysis, 2014. (Zitiert auf Seite 40)
- [387] Stiene, Stefan, Kai Lingemann, Andreas Nuchter und Joachim Hertzberg: *Contour-based object detection in range images*. In: *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, Seite 168–175, 2006. (Zitiert auf Seite 40)
- [388] Fod, A., A. Howard und M.J. Mataric: *Laser-based people tracking*. In: *IEEE International Conference on Robotics & Automation*, 2002. (Zitiert auf Seite 40)
- [389] Nakamura, Katsuyuki, Huijing Zhao, Ryosuke Shibasaki, Kiyoshi Sakamoto, Tomowo Ooga und Naoki Suzukawa: *Tracking pedestrian by using multiple laser range scanners*. Technical Report, Center for Spatial Information Science, University of Tokyo, 2004. (Zitiert auf Seite 40)
- [390] Zhao, H. und R. Shibasaki: *Pedestrian tracking using multiple laser range scanners*. In: *Proceedings of Computers on Urban Planning and Urban Management*, 2003. (Zitiert auf Seite 40)
- [391] Zhao, Huijing, Ryosuke Shibasaki und Nobuaki Ishihara: *Pedestrian Tracking using Single-row Laser Range Scanners*. In: *Machine Vision Applications (IAPR MVA), International Association for Pattern Recognition (IAPR) Conference on*, 2002. (Zitiert auf Seite 40)
- [392] Prassler, E., J. Scholz und A. Elfes: *Tracking people in a railway station during rush-hour*. Computer Vision Systems, Seite 162–179, 1999. (Zitiert auf Seite 40)
- [393] Spinello, Luciano, Rudolph Triebel und Roland Siegwart: *Multimodal people detection and tracking in crowded scenes*. In: *23rd national conference on Artificial intelligence*, Band 3, Seite 1409–1414, 2008. (Zitiert auf Seite 40)
- [394] Cui, Jinshi, Hongbin Zha, Huijing Zhao und Ryosuke Shibasaki: *Multi-modal tracking of people using laser scanners and video camera*. Image and Vision Computing, 26(2):240 – 252, 2008. (Zitiert auf Seite 40)
- [395] Cui, J., H. Zha, H. Zhao und R. Shibasaki: *Laser-based detection and tracking of multiple people in crowds*. Computer Vision and Image Understanding, 106(2-3):300–312, 2007. (Zitiert auf Seite 40)
- [396] Cui, Jinshi, Hongbin Zha, Huijing Zhao und R. Shibasaki: *Tracking multiple people using laser and vision*. In: *Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on*, Seite 2116–2121, 2005. (Zitiert auf Seite 40)
- [397] Murakita, Takuya, Tetsushi Ikeda und Hiroshi Ishiguro: *Human tracking using floor sensors based on the Markov chain Monte Carlo method*. In: *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, Band 4, Seite 917–920, 2004. (Zitiert auf Seite 41)
- [398] Valtonen, Miika, Jaakko Maentausta und Jukka Vanhala: *Tiletrack: Capacitive human tracking using floor tiles*. In: *Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on*, Seite 1–10, 2009. (Zitiert auf Seite 41)

- [399] Middleton, Lee, Alex A. Buss, Alex Bazin und Mark S. Nixon: *A floor sensor system for gait recognition*. In: *Automatic Identification Advanced Technologies, 2005. Fourth IEEE Workshop on*, Seite 171–176, 2005. (Zitiert auf Seite 41)
- [400] Audette, William E., David B. Kynor, Jed C. Wilbur, John R. Gagne, Lindamae Peck und N. H. Human: *Improved intruder detection using seismic sensors and adaptive noise cancellation*. In: *Human, Light Vehicle, and Tunnel Detection Workshop*, Seite 1–14, 2009. (Zitiert auf Seite 41)
- [401] Giannotti, Fosca, Mirco Nanni, Fabio Pinelli und Dino Pedreschi: *Trajectory pattern mining*. In: *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, Seite 330–339, 2007. (Zitiert auf Seite 41)
- [402] Varshavsky, Alex, Mike Y Chen, Eyal de Lara, Jon Froehlich, Dirk Haehnel, Jeffrey Hightower, Anthony LaMarca, Fred Potter, Timothy Sohn und Karen Tang: *Are GSM phones THE solution for localization?* In: *Mobile Computing Systems and Applications, 2006. WMCSA'06. Proceedings. 7th IEEE Workshop on*, Seite 34–42, 2005. (Zitiert auf Seite 41)
- [403] Drane, Christopher, Malcolm Macnaughtan und Craig Scott: *Positioning GSM telephones*. *Communications Magazine*, IEEE, 36(4):46–54, 1998. (Zitiert auf Seite 41)
- [404] Montjoye, Yves-Alexandre de, César A. Hidalgo, Michel Verleysen und Vincent D. Blondel: *Unique in the Crowd: The privacy bounds of human mobility*. *Scientific Reports*, 3, 2013. (Zitiert auf Seite 41)
- [405] Dobra, Adrian, Nathalie E. Williams und Nathan Eagle: *Spatiotemporal Detection of Unusual Human Population Behavior Using Mobile Phone Data*. *arXiv e-prints*, 2014. (Zitiert auf Seite 41)
- [406] Cheng, Yu-Chung, Yatin Chawathe, Anthony LaMarca und John Krumm: *Accuracy characterization for metropolitan-scale Wi-Fi localization*. In: *Proceedings of the 3rd international conference on Mobile systems, applications, and services*, Seite 233–245, 2005. (Zitiert auf Seite 42)
- [407] Teker, Ugur und Christian Freksa: *Realisierung und Evaluation eines Indoor-Lokalisierungssystems mittels WLAN*. *Dissertation*, Universität Bremen, 2005. (Zitiert auf Seite 42)
- [408] Kushki, Azadeh, Konstantinos N. Plataniotis und Anastasios N. Venetsanopoulos: *Intelligent dynamic radio tracking in indoor wireless local area networks*. *Mobile Computing, IEEE Transactions on*, 9(3):405–419, 2010. (Zitiert auf Seite 42)
- [409] Castro, Paul, Patrick Chiu, Ted Kremenek und Richard Muntz: *A probabilistic room location service for wireless networked environments*. In: *Ubiquitous Computing*, Seite 18–34, 2001. (Zitiert auf Seite 42)
- [410] Youssef, Moustafa und Ashok Agrawala: *The Horus location determination system*. *Wireless Networks*, 14(3):357–374, 2008. (Zitiert auf Seite 42)
- [411] Youssef, Moustafa und Ashok Agrawala: *The Horus WLAN location determination system*. In: *Proceedings of the 3rd international conference on Mobile systems, applications, and services*, Seite 205–218, 2005. (Zitiert auf Seite 42)

- [412] Kramer, André: *Orientierung im Innern - Lokalisierung und Navigation in öffentlichen Gebäuden*. c't magazin, 12:134–138, 2014. (Zitiert auf den Seiten 42 und 43)
- [413] Malinovskiy, Yegor, Nicolas Saunier und Yinhai Wang: *Analysis of pedestrian travel with static bluetooth sensors*. Transportation Research Record: Journal of the Transportation Research Board, 2299(1):137–149, 2012. (Zitiert auf Seite 42)
- [414] Cheung, Kenneth C., Stephen S. Intille und Kent Larson: *An inexpensive bluetooth-based indoor positioning hack*. Ubiquitous Computing, International Conference, 2006. (Zitiert auf Seite 42)
- [415] Faragher, Ramsey und Robert Harle: *An Analysis of the Accuracy of Bluetooth Low Energy for Indoor Positioning Applications*. In: *Proceedings of the 27th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2014)*, Seite 201–210, 2014. (Zitiert auf Seite 42)
- [416] Puscher, Frank: *Leuchtfener an der Decke - Wie Nahfunktechnik ortsbezogene Dienste aufs Smartphone bringt*. c't magazin, 12:140–143, 2014. (Zitiert auf Seite 42)
- [417] Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI): *Crowd-Monitoring makes the Olympics safer - Smartphone-apps support the security forces in London*, 2012. http://www.dfki.de/web/presse/pressemitteilungen_intern/2012/crowd%20monitoring (12.07.2014). (Zitiert auf Seite 42)
- [418] Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI): *Lang leve de Koning - Smartphone-App aus dem DFKI unterstützt die Sicherheitskräfte während der Krönung des niederländischen Königs*, 2013. http://www.dfki.de/web/presse/pressemitteilungen_intern/2013/201elang-leve-de-koning201c-2013-smartphone-app-aus-dem-dfki-unterstutzt-die-sicherheitskrafte-bei-der-kronung-des-niederlandischen-konigs/ (12.07.2014). (Zitiert auf Seite 42)
- [419] Kasemsuppakorn, Piyawan und Hassan A. Karimi: *A pedestrian network construction algorithm based on multiple {GPS} traces*. Transportation Research Part C: Emerging Technologies, 26(0):285 – 300, 2013. (Zitiert auf Seite 42)
- [420] Hillen, F., B. Höfle, M. Ehlers und P. Reinartz: *The potential of agent-based modelling for verification of people trajectories based on smartphone sensor data*. IOP Conference Series: Earth and Environmental Science, 18(1):012052, 2014. (Zitiert auf Seite 42)
- [421] Zheng, Yu, Lizhu Zhang, Xing Xie und Wei-Ying Ma: *Mining interesting locations and travel sequences from GPS trajectories*. In: *Proceedings of the 18th international conference on World wide web*, Seite 791–800, 2009. (Zitiert auf Seite 42)
- [422] Lymberopoulos, Dimitrios, Quentin Lindsey und Andreas Savvides: *An Empirical Characterization of Radio Signal Strength Variability in 3-D IEEE 802.15.4 Networks Using Monopole Antennas*. In: Römer, Kay, Holger Karl und Friedemann Mattern (Hrsg.): *Wireless Sensor Networks*, Band 3868 der Reihe *Lecture Notes in Computer Science*, Seite 326–341. Springer Berlin Heidelberg, 2006. (Zitiert auf Seite 42)

- [423] Oktem, Rusen und Elif Aydin: *An RFID based indoor tracking method for navigating visually impaired people*. Turk J Elec Eng & Comp Sci, 18(2):185–196, 2010. (Zitiert auf Seite 42)
- [424] Tabak, Vincent, Bauke de Vries und Jan Dijkstra: *RFID Technology Applied for Validation of an Office Simulation Model*. In: Klingsch, Wolfram W. F., Christian Rogsch, Andreas Schadschneider und Michael Schreckenberg (Hrsg.): *Pedestrian and Evacuation Dynamics 2008*, Seite 269–275, 2010. (Zitiert auf Seite 42)
- [425] Rodríguez-Moreno, María Dolores, Bonifacio Castaño, Melquiades Carbajo, Angel Moreno, David F. Barrero Moreno und Pablo Muñoz: *Multi-Agent Intelligent Planning Architecture for People Location and Orientation Using RFID*. Cybernetics and Systems, 42:16–32, 2011. (Zitiert auf Seite 42)
- [426] Ubisense: *Ubisense real-time location research products*. <http://ubisense.net/en/products-and-services/products/research-packages.html> (12.07.2014). (Zitiert auf Seite 42)
- [427] Domain, Time: *Time Domain ultra wideband (UWB) research products*. <http://www.timedomain.com/pulson-lab.php> (12.07.2014). (Zitiert auf Seite 42)
- [428] Haid, Markus: *Verbesserung der referenzlosen inertialen Objektverfolgung zur Low-cost Indoor-Navigation durch Anwendung der Kalman-Filterung*. Dissertation, Fraunhofer TEG, Stuttgart, 2005. (Zitiert auf Seite 43)
- [429] Krach, Bernhard und Patrick Robertson: *Integration of foot-mounted inertial sensors into a Bayesian location estimation framework*. In: *Positioning, Navigation and Communication, 2008. WPNC 2008. 5th Workshop on*, Seite 55–61, 2008. (Zitiert auf Seite 43)
- [430] Abdulrahim, Khairi, C. Hide, T. Moore und C. Hill: *Integrating low cost IMU with building heading in indoor pedestrian navigation*. Journal of Global Positioning Systems, 10(1):30–38, 2011. (Zitiert auf Seite 43)
- [431] Angermann, M. und P. Robertson: *FootSLAM: Pedestrian Simultaneous Localization and Mapping Without Exteroceptive Sensors – Hitchhiking on Human Perception and Cognition*. Proceedings of the IEEE, 100(Special Centennial Issue):1840–1848, 2012. (Zitiert auf Seite 43)
- [432] Godha, S. und G. Lachapelle: *Foot mounted inertial system for pedestrian navigation*. Measurement Science and Technology, 19(7):075202, 2008. (Zitiert auf Seite 43)
- [433] Jimenez, A.R., F. Seco, C. Prieto und J. Guevara: *A comparison of Pedestrian Dead-Reckoning algorithms using a low-cost MEMS IMU*. In: *Intelligent Signal Processing, 2009. WISP 2009. IEEE International Symposium on*, Seite 37–42, 2009. (Zitiert auf Seite 43)
- [434] Li, Jin-Feng, Qing-Hui Wang, Xiao-Mei Liu, Shun Cao und Feng-Long Liu: *A Pedestrian Dead Reckoning System Integrating Low-Cost MEMS Inertial Sensors and GPS Receiver*. Journal of Engineering Science and Technology Review, 7(2):197–203, 2014. (Zitiert auf Seite 43)
- [435] Jin, Yunye, Hong-Song Toh, Wee-Seng Soh und Wai-Choong Wong: *A Robust Dead-Reckoning Pedestrian Tracking System with Low Cost Sensors*. In: *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, Seite 222–230, 2011. (Zitiert auf Seite 43)

- [436] Ruiz, A. R. J., F. S. Granja, J. C. Prieto Honorato und J. I. G. Rosas: *Accurate Pedestrian Indoor Navigation by Tightly Coupling Foot-Mounted IMU and RFID Measurements*. Instrumentation and Measurement, IEEE Transactions on, 61(1):178–189, 2012. (Zitiert auf Seite 43)
- [437] Seitz, Jochen, Lucila Patiño-Studencka, Bernd Schindler, Stephan Haimmerl, Javier Gutiérrez Boronat, Steffen Meyer und Jörn Thielecke: *Sensor data fusion for pedestrian navigation using wlan and ins*. Technical Report, Fraunhofer Institute for Integrated Circuits, 2007. (Zitiert auf Seite 43)
- [438] Wang, Hui, Henning Lenz, Andrei Szabo, Joachim Bamberger und Uwe D Hanebeck: *WLAN-based pedestrian tracking using particle filters and low-cost MEMS sensors*. In: *Positioning, Navigation and Communication, 2007. WPNC'07. 4th Workshop on*, Seite 1–7, 2007. (Zitiert auf Seite 43)
- [439] Frank, Korbinian, Bernhard Krach, Noel Catterall und Patrick Robertson: *Development and evaluation of a combined wlan & inertial indoor pedestrian positioning system*. In: *ION GNSS*, 2009. (Zitiert auf Seite 43)
- [440] Foxlin, Eric: *Pedestrian tracking with shoe-mounted inertial sensors*. Computer Graphics and Applications, IEEE, 25(6):38–46, 2005. (Zitiert auf Seite 43)
- [441] Xsens: *3D motion tracking technology*. <http://www.xsens.com/> (12.03.2014). (Zitiert auf Seite 43)
- [442] Luo, Ren C und Ogst Chen: *Wireless and pyroelectric sensory fusion system for indoor human/robot localization and monitoring*. Mechatronics, IEEE/ASME Transactions on, 18(3):845–853, 2013. (Zitiert auf Seite 43)
- [443] Banerjee, Salil: *Improving Accuracy in Ultra-Wideband Indoor Position Tracking through Noise Modeling and Augmentation*. Dissertation, Clemson University, 2012. (Zitiert auf Seite 43)
- [444] Mandeljc, Rok, Stanislav Kovacic, Matej Kristan und Janez Pers: *Tracking by Identification Using Computer Vision and Radio*. Sensors, 13:241–273, 2013. (Zitiert auf Seite 43)
- [445] Fiedler, David Norbert: *Beiträge zur Analyse, Modellierung und Kalibrierung von Kameras und 3D-Tiefensensoren*. Dissertation, Technischen Universität Dortmund, 2014. (Zitiert auf Seite 45)
- [446] Bradski, G.: *The OpenCV Library*. Dr. Dobb's Journal of Software Tools, 2000. <http://opencv.org/> (25.11.2013). (Zitiert auf den Seiten 45, 91, 99 und 171)
- [447] Bradski, Gary R. und Vadim Pisarevsky: *Intel's Computer Vision Library: Applications in Calibration, Stereo, Segmentation, Tracking, Gesture, Face and Object Recognition*. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2:2796+, 2000. (Zitiert auf den Seiten 45, 91, 99 und 171)
- [448] Bouguet, J.-Y.: *Pyramidal Implementation of the Lucas Kanade Feature Tracker*. OpenCV Documents, 1999. (Zitiert auf den Seiten 45 und 70)
- [449] Lucas, Bruce D. und Takeo Kanade: *An Iterative Image Registration Technique with an Application to Stereo Vision*. In: *International Joint Conference on Artificial Intelligence*, Seite 674–679, 1981. (Zitiert auf Seite 45)

- [450] Tomasi, Carlo und Takeo Kanade: *Detection and Tracking of Point Features*. Technical Report CMU-CS-91-132, Carnegie Mellon University, 1991. (Zitiert auf Seite 45)
- [451] Shi, Jianbo und Carlo Tomasi: *Good Features to Track*. In: *IEEE Conference on Computer Vision and Pattern Recognition*, Seite 593–600, 1994. (Zitiert auf Seite 45)
- [452] Rameshbabu, K., J Swarnadurga, G Archana und K Menaka: *Target Tracking System Using Kalman Filter*. International Journal of Advanced Engineering Research and Studies, 2(1):90–94, 2012. (Zitiert auf den Seiten 45, 70 und 76)
- [453] Bradski, Gary R.: *Computer Vision Face Tracking For Use in a Perceptual User Interface*. Intel Technology Journal, 2:1–15, 1998. (Zitiert auf den Seiten 45 und 70)
- [454] Allen, John G., Richard Y. D. Xu und Jesse S. Jin: *Object tracking using CamShift algorithm and multiple quantized feature spaces*. In: *Proceedings of the Pan-Sydney area workshop on Visual information processing, VIP '05*, Seite 3–7, 2004. (Zitiert auf den Seiten 45 und 70)
- [455] Maggio, E. und A. Cavallaro: *Video Tracking: Theory and Practice*. Wiley, 2011. (Zitiert auf Seite 45)
- [456] Boltes, Maik: *PeTrack*. <http://www.fz-juelich.de/jsc/petrack/> (15.08.2014). (Zitiert auf den Seiten 48, 219, 290 und 292)
- [457] Schreer, Oliver: *Stereoanalyse und Bildsynthese*. Springer, 2005. (Zitiert auf den Seiten 51 und 89)
- [458] Salden, Daniel: *Dichtebestimmung von Menschenmassen unabhängig von der Kameraausrichtung*. Diplomarbeit, Fachhochschule Aachen Campus Jülich Labor für Medizinische Informatik, 2012. (Zitiert auf den Seiten 51 und 220)
- [459] Brown, D. C.: *Close-range camera calibration*. Photogrammetric Engineering, 37:855–866, 1971. (Zitiert auf den Seiten 53 und 121)
- [460] Zhang, Zhengyou: *A Flexible New Technique for Camera Calibration*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11):1330–1334, 2000. (Zitiert auf den Seiten 53 und 55)
- [461] Zhang, Zhengyou: *Flexible camera calibration by viewing a plane from unknown orientations*. In: *Proceeding of IEEE International Conference on Computer Vision (ICCV)*, Seite 666–673, 1999. (Zitiert auf Seite 55)
- [462] Sochman, Jan und David C. Hogg: *Who Knows Who - Inverting the Social Force Model for Finding Groups*. In: *Proceeding of IEEE International Conference on Computer Vision (ICCV) [630]*, Seite 830–837. (Zitiert auf Seite 60)
- [463] Kaewtrakulpong, P. und R. Bowden: *An Improved Adaptive Background Mixture Model for Realtime Tracking with Shadow Detection*. In: *2nd European Workshop on Advanced Video-Based Surveillance Systems*, 2001. (Zitiert auf Seite 61)
- [464] Fitzgibbon, Andrew W., Maurizio Pilu und Robert B. Fisher: *Direct least Square Fitting of Ellipses*. In: Pilu, M. und R. B. Fisher (Hrsg.): *Proceedings of the 13th International Conference on Pattern Recognition*, Band 1, Seite 253 – 257, 1996. (Zitiert auf den Seiten 62 und 109)

- [465] Wikipedia: *HSV-Farbraum*. <http://de.wikipedia.org/wiki/HSV-Farbraum> (12.05.2013). (Zitiert auf Seite 66)
- [466] Bender, Michael und Manfred Brill: *Computergrafik: Ein anwendungsorientiertes Lehrbuch*. Carl Hanser Verlag GmbH & Co, 2005. (Zitiert auf Seite 66)
- [467] Schaefer, C., L. Bergmann, H. J. Eichler und H. Niedrig: *Optik: Wellen- und Teilchenoptik*. Bergmann. Walter de Gruyter, 2004. (Zitiert auf Seite 66)
- [468] Watt, Alan: *3D Computer Graphics*. Addison-Wesley, 1999. (Zitiert auf Seite 66)
- [469] Gonzalez, Rafael C. und Richard E. Woods: *Digital Image Processing*. Prentice Hall, 2007. (Zitiert auf Seite 67)
- [470] Bradski, Gary R.: *Real time face and object tracking as a component of a perceptual user interface*. In: *Workshop on Applications of Computer Vision (WACV)*, Seite 214–219, 1998. (Zitiert auf Seite 70)
- [471] Adelson, E. H., C. H. Anderson, J. R. Bergen, P. J. Burt und J. M. Ogden: *Pyramid Methods in Image Processing*. In: *RCA Engineer*, Band 29, Kapitel 6, Seite 33–41. RCA Research and Engineering, 1984. (Zitiert auf Seite 70)
- [472] Seyfried, Armin, Maik Boltes, Jens Kähler, Wolfram Klingsch, Andrea Portz, Tobias Rupprecht, Andreas Schadschneider, Bernhard Steffen und Andreas Winkens: *Enhanced empirical data for the fundamental diagram and the flow through bottlenecks*. In: KLINGSCH, WOLFRAM W. F. et al. [633], Seite 145–156. (Zitiert auf den Seiten 72 und 206)
- [473] Arun, K. S., T. S. Huang und S. D. Blostein: *Least-squares fitting of two 3-D point sets*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 9:698–700, 1987. (Zitiert auf den Seiten 80 und 295)
- [474] Research, Point Grey: *Triclops*. <http://www.ptgrey.com/products/triclopssdk/> (25.11.2013). (Zitiert auf Seite 91)
- [475] Scharstein, D und R Szeliski: *A taxonomy and evaluation of dense two-frame stereo correspondence algorithms*. International Journal of Computer Vision, 47(1–3):7–42, 2002. (Zitiert auf den Seiten 91 und 95)
- [476] Murray, Don und James J. Little: *Using Real-Time Stereo Vision for Mobile Robot Navigation*. Autonomous Robots, 8(2):161–171, 2000. (Zitiert auf Seite 91)
- [477] Fua, Pascal: *Combining stereo and monocular information to compute dense depth maps that preserve depth discontinuities*. In: *In Proceedings of the 12th International Joint Conference on Artificial Intelligence*, Seite 1292–1298, 1991. (Zitiert auf Seite 91)
- [478] Konolige, Kurt: *Small Vision Systems: Hardware and Implementation*. In: Shirai, Yoshiaki und Shigeo Hirose (Hrsg.): *Robotics Research*, Seite 203–212. Springer London, 1998. (Zitiert auf Seite 91)
- [479] Konolige, Kurt und Motilal Agrawal: *Frame-frame matching for realtime consistent visual mapping*. In: *International Conference on Robotics and Automation (ICRA)*, 2007. (Zitiert auf Seite 91)

- [480] Hirschmuller, H.: *Accurate and efficient stereo processing by semi-global matching and mutual information*. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Band 2, Seite 807–814 vol. 2, 2005. (Zitiert auf Seite 91)
- [481] Birchfield, Stan und Carlo Tomasi: *Depth Discontinuities by Pixel-to-Pixel Stereo*. *International Journal of Computer Vision*, 35:1073–1080, 1996. (Zitiert auf Seite 91)
- [482] Schmid, Patricia: *Konstruktion eines Verbundes von Stereokameras zur kameraübergreifenden Personenverfolgung*. Diplomarbeit, Fachhochschule Aachen Campus Jülich, 2011. (Zitiert auf Seite 91)
- [483] Geiger, Andreas: *libelas - Library for Efficient Large-scale Stereo Matching*. <http://www.cvlibs.net/software/libelas/> (25.11.2013). (Zitiert auf Seite 92)
- [484] Geiger, Andreas, Martin Roser und Raquel Urtasun: *Efficient Large-Scale Stereo Matching*. In: *Asian Conference on Computer Vision (ACCV)*, 2010. (Zitiert auf Seite 92)
- [485] Heckenkamp, Christoph: *Stereo-Vision*. Technical Report 4, Inspect, 2012. (Zitiert auf Seite 95)
- [486] Jennings, J. A. M. und W. N. Charman: *Depth resolution in stereoscopic systems*. *Applied Optics*, 33(22):5192–5196, 1994. (Zitiert auf Seite 95)
- [487] Jähne, Bernd: *Digital Image Processing*. Springer Berlin Heidelberg, 2005. (Zitiert auf Seite 96)
- [488] Zhao, Wenyi und N. Nandhakumar: *Effects of camera alignment errors on stereoscopic depth estimates*. *Pattern Recognition*, 29(12):2115 – 2126, 1996. (Zitiert auf Seite 96)
- [489] Chang, C. und S. Chatterjee: *Quantization error analysis in stereo vision*. In: *Signals, Systems and Computers, 1992. 1992 Conference Record of The Twenty-Sixth Asilomar Conference on*, Seite 1037–1041 vol.2, 1992. (Zitiert auf Seite 96)
- [490] Ramakrishna, R.S. und B. Vaidyanathan: *Error analysis in stereo vision*. In: Chin, Roland und Ting-Chuen Pong (Hrsg.): *Computer Vision - ACCV'98*, Band 1351 der Reihe *Lecture Notes in Computer Science*, Seite 296–304. Springer Berlin Heidelberg, 1997. (Zitiert auf Seite 98)
- [491] Kytö, Mikko, Mikko Nuutinen und Pirkko Oittinen: *Method for measuring stereo camera depth accuracy based on stereoscopic vision*. In: *SPIE Proceedings, Measurement and Uncertainty II*, Band 7864, Seite 78640I–78640I–9, 2011. (Zitiert auf Seite 98)
- [492] Voigt, Andrea: *Körperbau, Gelenkbeweglichkeit und Handkräfte Erwachsener im Generationenvergleich*. Dissertation, Universität Potsdam, 2009. (Zitiert auf den Seiten 103, 115, 126 und 187)
- [493] Lange, Wolfgang und Armin Windel: *Kleine Ergonomische Datensammlung*. Technical Report, Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, 2005. (Zitiert auf den Seiten 103, 115 und 116)
- [494] Deutsches Institut für Normung e. V. (DIN): *DIN 33402-2:2005-12 Ergonomie - Körpermaße des Menschen - Teil 2: Werte*. Norm, Deutsches Institut für Normung e. V. (DIN), 2005. (Zitiert auf Seite 103)
- [495] Smith, Kevin, Daniel Gatica-Perez, Jean-Marc Odobez und Sileye Ba: *Evaluating multi-object tracking*. In: *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, Seite 36–36, 2005. (Zitiert auf Seite 120)

- [496] Wikipedia: *Beurteilung eines Klassifikators*. http://de.wikipedia.org/wiki/Beurteilung_eines_Klassifikators (12.05.2014). (Zitiert auf Seite 120)
- [497] Wikipedia: *Precision and recall*. http://en.wikipedia.org/wiki/Precision_and_recall (12.05.2014). (Zitiert auf Seite 120)
- [498] Research, Point Grey: *Stereo Accuracy and Error Modeling*. Technical Application Note TAN2004006, Point Grey Research, 2012. (Zitiert auf Seite 124)
- [499] Research, Point Grey: *Stereo Accuracy Chart*. Stereo Accuracy Chart (Microsoft Excel format), Point Grey Research, 2009. (Zitiert auf Seite 124)
- [500] Fang, Zhi-Ming, Wei-Guo Song, Xuan Liu, Wei Lv, Jian Ma und Xia Xiao: *A continuous distance model (CDM) for the single-file pedestrian movement considering step frequency and length*. *Physica A: Statistical Mechanics and its Applications*, 391(1):307 – 316, 2012. (Zitiert auf Seite 127)
- [501] Chen, Xin, Tianqi Yang und Jiaming Xu: *Cross-view gait recognition based on human walking trajectory*. *Journal of Visual Communication and Image Representation*, 25(8):1842–1855, 2014. (Zitiert auf Seite 127)
- [502] BenAbdelkader, C., R. Cutler und L. Davis: *Person identification using automatic height and stride estimation*. In: *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, Band 4, Seite 377–380 vol.4, 2002. (Zitiert auf Seite 127)
- [503] BenAbdelkader, C., R. Cutler und L. Davis: *Stride and cadence as a biometric in automatic person identification and verification*. In: *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, Seite 372–377, 2002. (Zitiert auf Seite 127)
- [504] BenAbdelkader, Chiraz, Ross Cutler und Larry Davis: *View-invariant Estimation of Height and Stride for Gait Recognition*. In: Tistarelli, Massimo, Josef Bigun und AnilK. Jain (Hrsg.): *Biometric Authentication*, Band 2359 der Reihe *Lecture Notes in Computer Science*, Seite 155–167. Springer Berlin Heidelberg, 2002. (Zitiert auf Seite 127)
- [505] Chraïbi, Mohcine: *Validated force-based modeling of pedestrian dynamics*. Hochschulschrift (Dissertation), Universität zu Köln, 2012. (Zitiert auf den Seiten 127 und 206)
- [506] Brüderlin, Markus (Herausgeber): *Die Kunst der Entschleunigung - Bewegung und Ruhe in der Kunst von Caspar David Friedrich bis Ai Weiwei*, Kapitel Aufnahme zur Erforschung der Fußgängerdynamik in einem Stadion, Seite 52. Hatje Cantz, Kunstmuseum Wolfsburg, 2012. (Zitiert auf Seite 130)
- [507] Dietrich, Stefan: *Einsam im Unglück*. *Frankfurter Allgemeine - Zeitung für Deutschland*, 172:1, 2010. (Zitiert auf Seite 130)
- [508] Chraïbi, Mohcine, Maik Boltes, Andreas Schadschneider und Armin Seyfried (Hrsg.): *Traffic and Granular Flow 2013*. Springer, 2015. (Zitiert auf den Seiten 130 und 206)
- [509] Lossau, Norbert: *Chirurgen suchen nach der optimalen Kopfform*. *Die Welt*, 2011. (Zitiert auf Seite 135)

- [510] Waldeyer, Anton und Anton Mayet: *Anatomie des Menschen / Kopf und Hals, Auge, Ohr, Gehirn, Arm, Brust*. De Gruyter, 1965. (Zitiert auf Seite 135)
- [511] Bammes, Gottfried: *Die Gestalt des Menschen: Lehr- und Handbuch der Künstleranatomie*. Englisch Verlag, 2009. (Zitiert auf Seite 135)
- [512] University, Carnegie Mellon: *Panda3D*. <https://www.panda3d.org/> (15.01.2014). (Zitiert auf Seite 141)
- [513] Cognimatics: *TrueView People Counter*. <http://www.cognimatics.com/products/people-counter/> (15.03.2014). (Zitiert auf den Seiten 163 und 164)
- [514] Vitracom: *Shop Efficiency Monitoring*. <http://www.vitracom.de/> (15.03.2014). (Zitiert auf Seite 163)
- [515] nvl²: *MoPeCo*. <http://www.nvlhoch2.de/> (15.03.2014). (Zitiert auf den Seiten 163 und 164)
- [516] Desoi, Monika und Silke Jandt: *Zulässige Erhebung von Daten zu Forschungszwecken*. Datenschutz und Datensicherheit, 12:895–901, 2012. (Zitiert auf den Seiten 169 und 223)
- [517] Kim, S. J., D. Gallup, J. m. Frahm, A. Akbarzadeh, Q. Yang, R. Yang, D. Nistér und M. Pollefeys: *Gain adaptive real-time stereo streaming*. In: *Int. Conf. on Vision Systems*, 2007. (Zitiert auf Seite 171)
- [518] Stam, Joe: *Stereo imaging with cuda*. OpenVIDIA, Seite 1–20, 2008. (Zitiert auf Seite 171)
- [519] Gallup, David, Jan-Michael Frahm und Joe Stam: *Real-Time Local Stereo using CUDA*. In: *Proceedings of Nvidia GPU Technology Conference Research Summit*, 2009. (Zitiert auf Seite 171)
- [520] Banz, C., H. Blume und P. Pirsch: *Real-time semi-global matching disparity estimation on the GPU*. In: *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, Seite 514–521, 2011. (Zitiert auf Seite 171)
- [521] Wojek, Christian, Gyuri Dorko, Andre Schulz und Bernt Schiele: *Sliding-windows for rapid object class localization: A parallel technique*. In: *Proceedings of the 30th DAGM symposium on Pattern Recognition*, Seite 71–81, 2008. (Zitiert auf Seite 171)
- [522] Prisacariu, Victor und Ian Reid: *fastHOG - a real-time GPU implementation of HOG*. Technical Report 2310/09, Department of Engineering Science, Oxford University, 2009. (Zitiert auf Seite 171)
- [523] Petersen, Lizette: *Pedestrian experiment to verify a Hopf bifurcation in pedestrian flow*. DTU newspaper, 2012. (Zitiert auf Seite 172)
- [524] Jensen, Tore Vind: *Matematikere sætter trængsel på formel*. DTU Avisen, 7:16–17, 2012. (Zitiert auf Seite 172)
- [525] 1873/1-1, DFG Project KL und SE 1789/1-1: *Boundary Effects and Non-Equilibrium States in Pedestrian Dynamics - Experiments and Modelling*. http://www.fz-juelich.de/ias/jsc/EN/Research/ModellingSimulation/CivilSecurityTraffic/Projects/DFG/_node.html (12.03.2013). (Zitiert auf Seite 182)

- [526] Klüpfel, Hubert, Armin Seyfried, Stefan Holl, Maik Boltes, Mohcine Chraïbi, Ulrich Kemloh, Andrea Portz, Jack Little, Tobias Rupprecht, Andreas Winkens, Wolfram Klingsch, Christian Eilhardt, Stefan Nowak, Andreas Schadschneider, Tobias Kretz und Michael Krabbe: *HERMES - Evacuation Assistant for Arenas*. In: *Future Security*, 2010. (Zitiert auf den Seiten 182, 183 und 206)
- [527] Holl, Stefan, Armin Seyfried und Maik Boltes: *Hermes - Erforschung eines Evakuierungsassistenten für den Krisenfall bei Großveranstaltungen*. In: Fähnrich, K.P. und B. Franczyk (Hrsg.): *IT-Unterstützung von Rettungskräften, Informatik 2010, Service Science - Neue Perspektiven für die Informatik*, Band 2, Seite 162–167, 2010. (Zitiert auf den Seiten 182, 183, 206, 211 und 215)
- [528] BaSiGo: *Bausteine für die Sicherheit von Großveranstaltungen*. <http://www.basigo.de> (12.03.2013). (Zitiert auf Seite 182)
- [529] Lee, Jodie Y. S. und William H. K. Lam: *The Variation of Walking Speeds on a Uni-directional Walkway and on a Bi-directional Stairway*. *Transportation Research Record*, 1982:122–131, 2006. (Zitiert auf Seite 181)
- [530] LI Xiang, DONG Li-Yun: *Modeling and Simulation of Pedestrian Counter Flow on a Crosswalk*. *Chinese Physics Letters*, 29(9):98902, 2012. (Zitiert auf Seite 181)
- [531] Yanagisawa, Daichi, Akiyasu Tomoeda und Katsuhiro Nishinari: *Improvement of pedestrian flow by slow rhythm*. *Physical Review E*, 85:016111+, 2012. (Zitiert auf Seite 181)
- [532] Burghardt, Sebastian: *Beitrag zur Optimierung der Personenführung in Stadien*. Hochschulschrift (Dissertation), Bergische Universität Wuppertal, 2013. (Zitiert auf den Seiten 183, 206, 207 und 213)
- [533] Forschungszentrum Jülich, Jülich Supercomputing Centre: *Database of own experiments*. <http://ped.fz-juelich.de/database> (15.07.2014). (Zitiert auf den Seiten 183, 207 und 219)
- [534] Allain, Pierre, Nicolas Courty und Thomas Corpetti: *AGORASET: a dataset for crowd video analysis*. In: *1st ICPR International Workshop on Pattern Recognition and Crowd Analysis*, Seite 1–6, 2012. (Zitiert auf Seite 183)
- [535] Seyfried, Armin, Bernhard Steffen, Wolfram Klingsch und Maik Boltes: *The fundamental diagram of pedestrian movement revisited*. *Journal of Statistical Mechanics: Theory and Experiment*, P10002, 2005. (Zitiert auf den Seiten 187 und 206)
- [536] Chattaraj, Ujjal, Armin Seyfried und Partha Chakroborty: *Comparison of Pedestrian Fundamental Diagram Across Cultures*. *Advances in Complex Systems (ACS)*, 12(3):393–405, 2009. (Zitiert auf den Seiten 187 und 206)
- [537] Lukowski, Janine: *Einfluss der Motivation auf das Fundamentaldiagramm der Fußgängerbewegung entlang einer Linie*. Masterthesis, Bergische Universität Wuppertal, 2009. (Zitiert auf den Seiten 187 und 206)
- [538] dpreview: *Making (some) sense out of sensor sizes*. <http://www.dpreview.com/news/2002/10/7/sensorsizes> (12.04.2013). (Zitiert auf den Seiten 192 und 194)
- [539] IDS: *UI-2230-C*. <http://de.ids-imaging.com/store/ui-2230se.html> (16.11.2013). (Zitiert auf Seite 194)

- [540] IDS: *UI-3240CP-M*. <http://de.ids-imaging.com/store/ui-3240cp.html> (16.11.2013). (Zitiert auf Seite 194)
- [541] Research, Point Grey: *Bumblebee*. <http://www.ptgrey.com/products/bumblebee/bumblebee.PDF> (15.03.2013). (Zitiert auf Seite 194)
- [542] Research, Point Grey: *Bumblebee XB3*. <http://www.ptgrey.com/products/bbxb3/> (15.03.2013). (Zitiert auf Seite 194)
- [543] Sony: *DCR-VX2000E*. <http://www.sony.de/support/de/product/dcr-vx2000e> (16.07.2014). (Zitiert auf Seite 194)
- [544] Sony: *HVR-Z5E*. <http://www.sony.de/pro/product/broadcast-products-camcorders-hdv-dvcam/hvr-ale/overview> (9.11.2013). (Zitiert auf Seite 194)
- [545] Sony: *CX740VE*. <http://www.sony.de/support/de/product/hdr-cx740ve> (06.07.2014). (Zitiert auf Seite 194)
- [546] Instruments, National: *Auswahl der geeigneten Keraschnittstelle für Ihr Bildverarbeitungssystem*. Technical Report, National Instruments, 2007. (Zitiert auf Seite 195)
- [547] Research, Point Grey: *MultiSync*. <http://www.ptgrey.com/products/multisync/> (25.11.2013). (Zitiert auf Seite 196)
- [548] Research, Point Grey: *Extending the Distance between a Camera and the Controlling Host System*. Technical Application Note TAN2008009, Point Grey Research, 2012. (Zitiert auf Seite 197)
- [549] Wikipedia: *Datendurchsatz*. <http://de.wikipedia.org/wiki/Datendurchsatz> (12.03.2013). (Zitiert auf Seite 198)
- [550] Adobe: *After Effects*. <http://www.adobe.com/de/products/aftereffects.html> (15.04.2013). (Zitiert auf Seite 201)
- [551] Steffen, Bernhard, Armin Seyfried und Maik Boltes: *Reliability issues in the microscopic modeling of pedestrian movement*. In: Pavel Exner, Hradec Kralove (Herausgeber): *Mathematical results in Quantum Physics*, Seite 254–259, 2011. (Zitiert auf den Seiten 203 und 206)
- [552] Steffen, Bernhard und Armin Seyfried: *Methods for measuring pedestrian density, flow, speed and direction with minimal scatter*. *Physica A*, 389(9):1902–1910, 2010. (Zitiert auf den Seiten 203, 204 und 206)
- [553] Voronoi, G. M.: *Nouvelles applications des paramètres continus à la théorie des formes quadratiques*. *Journal für die reine und angewandte Mathematik*, 133:198–287, 1908. (Zitiert auf Seite 204)
- [554] Zhang, Jun: *Pedestrian fundamental diagrams: Comparative analysis of experiments in different geometries*. Dissertation, Bergische Universität Wuppertal, 2012. (Zitiert auf den Seiten 205, 206 und 207)
- [555] Seyfried, Armin, Bernhard Steffen, Wolfram Klingsch, Thomas Lippert und Maik Boltes: *Steps Toward the Fundamental Diagram - Empirical Results and Modelling*. In: WALDAU, N. et al. [628], Seite 377–390. (Zitiert auf Seite 206)

- [556] Seyfried, Armin, Bernhard Steffen, Wolfram Klingsch, Thomas Lippert und Maik Boltes: *The Fundamental Diagram of Pedestrian Movement Revisited - Empirical Results and Modelling*. In: Schadschneider, Andreas, T. Pöschel, R. Kühne, Michael Schreckenberg und Dieter E. Wolf (Hrsg.): *Traffic and Granular Flow 2005*, Seite 305–314, 739, 2007. (Zitiert auf Seite 206)
- [557] Chattaraj, Ujjal, Armin Seyfried und Partha Chakroborty: *Modeling Single File Pedestrian Motion Across Cultures*. Safety Science, 2009. (Zitiert auf Seite 206)
- [558] Portz, Andrea und Armin Seyfried: *Modeling Stop-and-Go Waves in Pedestrian Dynamics*. In: Wyrzykowski, Roman, Jack Dongarra, Konrad Karczewski und Jerzy Wasniewski (Hrsg.): *Parallel Processing and Applied Mathematics 2009, Part II*, Seite 561–568, 2010. (Zitiert auf Seite 206)
- [559] Seyfried, Armin, Bernhard Steffen und Thomas Lippert: *Basics of modelling the pedestrian flow*. Physica A, 368:232–238, 2006. (Zitiert auf Seite 206)
- [560] Portz, Andrea und Armin Seyfried: *Analyzing Stop-and-Go Waves by Experiment and Modeling*. In: PEACOCK, RICHARD D. et al. [632], Seite 577–586. (Zitiert auf Seite 206)
- [561] Schadschneider, Andreas und Armin Seyfried: *Empirical Results for Pedestrian Dynamics and their Implications for Cellular Automata Models*. In: Timmermans, Harry (Herausgeber): *Pedestrian Behavior: Data Collection and Applications*, Kapitel 2, Seite 27–43. Emerald Group Publishing Limited, 1 Auflage, 2009. (Zitiert auf Seite 206)
- [562] Schadschneider, Andreas, Hubert Klüpfel, Tobias Kretz, Christian Rogsch und Armin Seyfried: *Fundamentals of Pedestrian and Evacuation Dynamics*. In: Bazzan, Ana und Franziska Klügl (Hrsg.): *Multi-Agent Systems for Traffic and Transportation Engineering*, Kapitel 6, Seite 124–154. IGI Global, Hershey, Pennsylvania, USA, 2009. (Zitiert auf Seite 206)
- [563] Steffen, Bernhard und Armin Seyfried: *The repulsive force in continous space models of pedestrian movement*. arXiv e-prints, 2008. <http://arxiv.org/abs/0803.1319> (10.07.2014). (Zitiert auf Seite 206)
- [564] Schadschneider, Andreas und Armin Seyfried: *Validation of CA Models of pedestrian dynamics with fundamental diagrams*. Cybernetics and Systems, 40(5):367–389, 2009. (Zitiert auf den Seiten 206 und 215)
- [565] Rupprecht, Tobias, Armin Seyfried, Wolfram Klingsch und Maik Boltes: *Bottleneck capacity estimation for pedestrian traffic*. In: *Interflam 2007* [629], Seite 1423–1430. (Zitiert auf Seite 206)
- [566] Seyfried, Armin, Tobias Rupprecht, Andreas Winkens, Oliver Passon, Bernhard Steffen, Wolfram Klingsch und Maik Boltes: *Capacity Estimation for emergency exits and bottleneck*. In: *Interflam 2007* [629], Seite 247–258. (Zitiert auf Seite 206)
- [567] Seyfried, Armin, Bernhard Steffen, Andreas Winkens, Tobias Rupprecht, Maik Boltes und Wolfram Klingsch: *Empirical data for pedestrian flow through bottlenecks*. In: Appert-Rolland, Cécile, Francois Chevoir, Philippe Gondret, Sylvain Lassarre, Jean-Patrick Lebacque und Michael Schreckenberg (Hrsg.): *Traffic and Granular Flow '07*, Seite 189–199, 2009. (Zitiert auf Seite 206)

- [568] Winkens, Andreas, Tobias Rupprecht, Armin Seyfried und Wolfram Klingsch: *Empirical Study of Pedestrians' Characteristics at Bottlenecks*. In: Klingsch, Wolfram W. F., Christian Rogsch, Andreas Schadschneider und Michael Schreckenberg (Hrsg.): *Pedestrian and Evacuation Dynamics 2008*, Seite 263–268, 2010. (Zitiert auf Seite 206)
- [569] Chraïbi, Mohcine, Ulrich Kemloh, Armin Seyfried und Andreas Schadschneider: *Force-based models of pedestrian dynamics*. *Networks and Heterogeneous Media*, 6(3):425–442, 2011. (Zitiert auf den Seiten 206 und 215)
- [570] Holl, Stefan und Armin Seyfried: *Wissenschaftliche Hintergründe zur Bemessung von Rettungswegen - Ergebnisse aktueller Forschung*. Ingenieurakademie West e.V. (Hrsg.), Brandschutztagung 2011, 2011. (Zitiert auf Seite 206)
- [571] Boltes, Maik, Armin Seyfried, Bernd Steffen und Andreas Schadschneider: *Automatic Extraction of Pedestrian Trajectories from Video Recordings*. In: KLINGSCH, WOLFRAM W. F. et al. [633], Seite 43–54. (Zitiert auf Seite 206)
- [572] Boltes, Maik, Mohcine Chraïbi, Stefan Holl, Armel Ulrich Kemloh Wagoum, Gregor Lämmel, Weichen Liao, Wolfgang Mehner, Antoine Tordeux und Jun Zhang: *Experimentation, data collection, modeling and simulation of pedestrian dynamics*. In: *Statistics, probability and numerical analysis 2014*, 2014. (Zitiert auf Seite 206)
- [573] Winkens, Andreas, Wolfram Klingsch und Armin Seyfried: *New Data for Human Performance in Planar Corridors*. In: PEACOCK, RICHARD D. et al. [632], Seite 61–70. (Zitiert auf Seite 206)
- [574] Seyfried, Armin, Mohcine Chraïbi, Jonas Mehlich und Andreas Schadschneider: *Runtime Optimization of Force Based Models within the Hermes Project*. In: PEACOCK, RICHARD D. et al. [632], Seite 363–374. (Zitiert auf Seite 206)
- [575] Schadschneider, Andreas und Armin Seyfried: *Modeling pedestrian dynamics - From experiment to theory and back*. In: *Traffic and Granular Flow '09*, 2009. (Zitiert auf Seite 206)
- [576] Schadschneider, Andreas und Armin Seyfried: *Empirical results for pedestrian dynamics and their implications for modeling*. *Networks and Heterogeneous Media*, 6(3):545–560, 2011. (Zitiert auf den Seiten 206 und 208)
- [577] Seyfried, Armin und Andreas Schadschneider: *Fundamental Diagram and Validation of Crowd Models*. In: Umeo, H., S. Morishita, K. Nishinari, T. Komatsuzaki und S. Bandini (Hrsg.): *Cellular Automata*, Band 5191/2008 der Reihe *Lecture Notes in Computer Science*, Seite 563–566, 2008. (Zitiert auf Seite 206)
- [578] Seyfried, Armin und Andreas Schadschneider: *Validation of cellular automata models of pedestrian dynamics using controlled large-scale experiments*. In: *Cybernetics and Systems 2008*, 2008. (Zitiert auf Seite 206)
- [579] Liddle, Jack, Armin Seyfried, Wolfram Klingsch, Tobias Rupprecht, Andreas Schadschneider und Andreas Winkens: *An Experimental Study of Pedestrian Congestions: Influence of Bottleneck Width and Length*. In: *Traffic and Granular Flow 2009*, 2009. (Zitiert auf den Seiten 206 und 209)
- [580] Liddle, Jack, Armin Seyfried und Bernhard Steffen: *Analysis of bottleneck motion using Voronoi diagrams*. In: PEACOCK, RICHARD D. et al. [632], Seite 833–836. (Zitiert auf Seite 206)

- [581] Rupprecht, Tobias, Wolfram Klingsch und Armin Seyfried: *Influence of Geometry Parameters on Pedestrian Flow through Bottleneck*. In: PEACOCK, R.D. et al. [632], Seite 71–80. (Zitiert auf Seite 206)
- [582] Seyfried, Armin und Andreas Schadschneider: *Empirical Results for Pedestrian Dynamics at Bottlenecks*. In: Wyrzykowski, Roman, Jack Dongarra, Konrad Karczewski und Jerzy Wasniewski (Hrsg.): *Parallel Processing and Applied Mathematics 2009, Part II*, Seite 575–584, 2010. (Zitiert auf Seite 206)
- [583] Steffen, Bernhard, Armin Seyfried und Maik Boltes: *Reliability issues in the microscopic modeling of pedestrian movement*. In: *QMath 11 - Mathematical Results in Quantum Physics, WSPC*, 2010. (Zitiert auf Seite 206)
- [584] Chraïbi, Mohcine, Martina Freialdenhoven, Andreas Schadschneider und Armin Seyfried: *Modeling the desired direction in a force-based model for pedestrian dynamics*. In: *Traffic and Granular Flow 2011*, to appear. (Zitiert auf den Seiten 206 und 215)
- [585] Steffen, Bernhard, Maik Boltes und Armin Seyfried: *Improved Methods for Checking Forces Based Models of Pedestrian Dynamics*. In: PEACOCK, RICHARD D. et al. [632], Seite 885–888. (Zitiert auf Seite 206)
- [586] Chraïbi, Mohcine, Armin Seyfried, Andreas Schadschneider und Wolfgang Mackens: *Quantitative Description of Pedestrian Dynamics with a Force-based Model*. In: *IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology 2009*, Band 3, Seite 583–586, 2009. (Zitiert auf Seite 206)
- [587] Chraïbi, Mohcine, Armin Seyfried, Andreas Schadschneider und Wolfgang Mackens: *Quantitative Verification of a Force-based Model for Pedestrian Dynamics*. In: *Traffic and Granular Flow 2009*, 2009. (Zitiert auf Seite 206)
- [588] Boltes, Maik, Armin Seyfried, Bernhard Steffen und Andreas Schadschneider: *Using Stereo Recordings to Extract Pedestrian Trajectories Automatically in Space*. In: PEACOCK, RICHARD D. et al. [632], Seite 751–754. (Zitiert auf Seite 206)
- [589] Boltes, Maik, Mohcine Chraïbi, Ulrich Kemloh und Jack Little: *Tracking, Analysis and Simulation of Crowds*, 2010. ASIM-Workshop - Trends in Computational Science and Engineering: Foundations of Modeling and Simulation. (Zitiert auf Seite 206)
- [590] Burghardt, Sebastian, Armin Seyfried und Wolfram Klingsch: *Improving Egress Design through Measurement and Correct Interpretation of the Fundamental Diagram for Stairs*. In: Panda, Mahabir und Ujjal Chattaraj (Hrsg.): *Developments in Road Transportation*, Seite 181–187, 2010. (Zitiert auf den Seiten 206 und 207)
- [591] Holl, Stefan und Armin Seyfried: *Computersimulationen zur Analyse und Optimierung des Evakuierungsprozesses*. sicher ist sicher. Arbeitsschutz aktuell, 61:368–371, 2010. (Zitiert auf den Seiten 206 und 215)
- [592] Holl, Stefan und Armin Seyfried: *Validität von Evakuierungssimulationen*. vfdb-Zeitschrift, 1:35–41, 2010. (Zitiert auf den Seiten 206 und 216)

- [593] Zhang, Jun, Wolfram Klingsch und Armin Seyfried: *High precision analysis of unidirectional pedestrian flow within the Hermes Project*. In: *Fifth Performance-based Fire Protection and Fire Protection Engineering Seminars*, 2011. (Zitiert auf Seite 206)
- [594] Zhang, Jun, Wolfram Klingsch, Andreas Schadschneider und Armin Seyfried: *Transitions in pedestrian fundamental diagrams of straight corridors and T-junctions*. *Journal of Statistical Mechanics: Theory and Experiment*, P06004, 2011. (Zitiert auf den Seiten 206, 207 und 214)
- [595] Zhang, Jun und Armin Seyfried: *Empirical characteristics of different types of pedestrian streams*. In: *9th Asia-Oceania Symposium on Fire Science and Technology*, 2012. (Zitiert auf den Seiten 206 und 207)
- [596] Liao, Weichen, Armin Seyfried, Jun Zhang, Maik Boltes, Xiaoping Zheng und Ying Zhao: *Experimental study on pedestrian flow through wide bottleneck*. In: HOOGENDOORN, SERGE et al. [634]. (Zitiert auf den Seiten 206 und 209)
- [597] Chraïbi, Mohcine und Armin Seyfried: *Pedestrian Dynamics With Event-driven Simulation*. In: KLINGSCH, WOLFRAM W. F. et al. [633], Seite 713–718. (Zitiert auf Seite 206)
- [598] Zhang, Jun, Wolfram Klingsch, Andreas Schadschneider und Armin Seyfried: *Experimental study of pedestrian flow through a T-junction*. In: *Traffic and Granular Flow 2011*, 2011. (Zitiert auf den Seiten 206, 207 und 214)
- [599] Zhang, Jun, Wolfram Klingsch, Tobias Rupprecht, Andreas Schadschneider und Armin Seyfried: *Empirical study of turning and merging of pedestrian streams in T-junction*. In: *Fourth International Symposium on Agent-Based Modeling and Simulation*, 2012. (Zitiert auf den Seiten 206, 207, 212 und 214)
- [600] Zhang, Jun, Wolfgang Mehner, Erik Andresen, Stefan Holl, Maik Boltes, Andreas Schadschneider und Armin Seyfried: *Comparative analysis of pedestrian, bicycle and car traffic moving in circuits*. In: *2nd Conference of Transportation Research Group of India (2nd CTRG)*, 2013. (Zitiert auf den Seiten 206 und 207)
- [601] Seyfried, Armin, Erik Andresen, Maik Boltes, Stefan Holl, Wolfgang Mehner, Andreas Schadschneider und Jun Zhang: *Universalities in fundamental diagrams of cars, bicycles and pedestrians*. In: *Summer Symposium Celebrating 50 Years of Traffic Flow Theory*, 2014. (Zitiert auf Seite 206)
- [602] Zhang, Jun, Wolfgang Mehner, Stefan Holl, Maik Boltes, Erik Andresen, Andreas Schadschneider und Armin Seyfried: *Universal flow-density relation of single-file bicycle, pedestrian and car motion*. *Physics Letters A*, 378(44):3274 – 3277, 2014. (Zitiert auf Seite 206)
- [603] Tomoeda, Akiyasu, Daichi Yanagisawa und Katsuhiro Nishinari: *Escape Velocity of the Leader in a Queue of Pedestrians*. In: *Traffic and Granular Flow 2013*, 2015. (Zitiert auf Seite 206)
- [604] Ezaki, Takahiro, Kazumichi Ohtsuka, Daichi Yanagisawa und Katsuhiro Nishinari: *Inflow process: a counterpart of evacuation*. In: *Traffic and Granular Flow 2013*, 2015. (Zitiert auf Seite 206)
- [605] Fang, Jun, Zheng Qin, Hao Hu, Zhaohui Xu und Huan Li: *The fundamental diagram of pedestrian model with slow reaction*. *Physica A: Statistical Mechanics and its Applications*, 391(23):6112 – 6120, 2012. (Zitiert auf Seite 207)

- [606] Fang, Jun, Zheng Qin, Zhengcai Lu und Fengfei Zhao: *Fundamental diagram of pedestrian dynamics by safety interspace model*. arXiv e-prints, 2013. (Zitiert auf Seite 207)
- [607] Müller, Klaus: *Zur Gestaltung und Bemessung von Fluchtwegen für die Evakuierung von Personen aus Bauwerken auf der Grundlage von Modellversuchen*. Dissertation, Technische Hochschule Magdeburg, 1981. (Zitiert auf Seite 208)
- [608] Holl, Stefan und Armin Seyfried: *Hermes - an Evacuation Assistant for Mass Events*. inSiDe, 7(1):60–61, 2009. (Zitiert auf den Seiten 211 und 215)
- [609] Korhonen, Timo und Simo Hostikka: *Fire Dynamics Simulator with Evacuation: FDS+Evac - Technical Reference and User's Guide*. Technical Report, VTT Technical Research Centre of Finland, 2010. (Zitiert auf den Seiten 212 und 213)
- [610] Kemloh Wagoum, Armel Ulrich, Mohcine Chraïbi, Christian Eilhardt, Stefan Nowak, Igor Kulikov, Daniel Weber, Kathrin Sauer, Hubert Klüpfel und Andreas Schadschneider: *OpenPedSim: A framework for pedestrian flow analysis*. In: Weidmann, Ulrich, Uwe Kirsch und Michael Schreckenberg (Hrsg.): *Pedestrian and Evacuation Dynamics 2012*, Seite 1323–1330, 2014. (Zitiert auf Seite 214)
- [611] Kemloh Wagoum, Armel Ulrich und Armin Seyfried: *Optimizing the evacuation time of pedestrians in a graph-based navigation*. In: Panda, Mahabir und Ujjal Chattararaj (Hrsg.): *In Developments in Road Transportation*, Seite 188–196, 2010. (Zitiert auf Seite 216)
- [612] Wagoum, Armel Ulrich Kemloh, Bernhard Steffen, Armin Seyfried und Mohcine Chraïbi: *Parallel real time computation of large scale pedestrian evacuations*. elsevier, 2012. (Zitiert auf Seite 216)
- [613] Kemloh Wagoum, Armel Ulrich, Armin Seyfried und Stefan Holl: *Modelling dynamic route choice of pedestrians to assess the criticality of building evacuation*. *Advances in Complex Systems*, 15(3), 2012. (Zitiert auf Seite 216)
- [614] Kemloh Wagoum, Armel Ulrich: *Route choice modelling and runtime optimisation for simulation of building evacuation*. Dr., Bergische Universität Wuppertal, Jülich, 2013. Univ., Diss., 2012. (Zitiert auf Seite 216)
- [615] Kemloh Wagoum, Armel Ulrich, Bernhard Steffen und Armin Seyfried: *Runtime Optimisation Approaches for a Real-Time Evacuation Assistant*. In: Wyrzykowski, R. (Herausgeber): *9th Conference on Parallel Processing and Applied Mathematics*, Band 7203 der Reihe *Lecture Notes in Computer Science*, Seite 386–395, 2012. (Zitiert auf Seite 216)
- [616] Schadschneider, Andreas, Christian Eilhardt, Stefan Nowak und Robinson Will: *Towards a Calibration of the Floor Field Cellular Automaton*. In: Peacock, Richard D., Erica D. Kuligowski und Jason D. Averill (Hrsg.): *Pedestrian and Evacuation Dynamics*, Seite 557–566, 2011. (Zitiert auf Seite 215)
- [617] Wu, Yihong und Zhanyi Hu: *PnP problem revisited*. *Journal of Mathematical Imaging and Vision*, 24(1):131–141, 2006. (Zitiert auf Seite 220)
- [618] Guo, Ren-Yong, Hai-Jun Huang und S. C. Wong: *Route choice in pedestrian evacuation under conditions of good and zero visibility: Experimental and simulation results*. *Transportation Research Part B: Methodological*, 46(6):669 – 686, 2012. (Zitiert auf Seite 221)

- [619] Isobe, Motonari, Dirk Helbing und Takashi Nagatani: *Experiment, theory, and simulation of the evacuation of a room without visibility*. Physical Review E, 69:066132, 2004. (Zitiert auf Seite 221)
- [620] Zuriguel, Iker, Daniel Ricardo Parisi, Raúl Cruz Hidalgo, Celia Lozano, Alvaro Janda, Paula Alejandra Gago, Juan Pablo Peralta, Luis Miguel Ferrer, Luis Ariel Pugnaroni, Eric Clément, Diego Maza, Ignacio Pagonabarraga und Angel Garcimartín: *Clogging transition of many-particle systems flowing through bottlenecks*. Scientific Reports, 4(7324):1–8, 2014. (Zitiert auf Seite 222)
- [621] Williams, R. M. (Herausgeber): *OUTBACK Calendar*. Outback Publishing Company, 2013. (Zitiert auf Seite 222)
- [622] John, Alexander, Andreas Schadschneider, Debashish Chowdhury und Katsuhiro Nishinari: *Trafficklike Collective Movement of Ants on Trails: Absence of a Jammed Phase*. Physical Review Letters, 102:108001, 2009. (Zitiert auf Seite 222)
- [623] Shafique, Khurram und Mubarak Shah: *A noniterative greedy algorithm for multiframe point correspondence*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 27(1):51–65, 2005. (Zitiert auf Seite 223)
- [624] Coppack, Timothy, Christoph Kulemeyer und Axel Schulz: *Monitoring migratory birds through fixed pencil beam radar and infrared videography at offshore wind farm alpha ventus*. Environmental research at alpha ventus - first results: migratory birds, Seite 63–64, 2010. (Zitiert auf Seite 223)
- [625] Roussel, J.-M., A. Haro und R. A. Cunjak: *Field test of a new method for tracking small fishes in shallow rivers using passive integrated transponder (PIT) technology*. Canadian Journal of Fisheries and Aquatic Sciences, 57(7):1326–1329, 2000. (Zitiert auf Seite 223)
- [626] Correll, Nikolaus, Gregory Sempo, Yuri Lopez de Meneses, Jose Halloy, Jean-Louis Deneubourg und Alcherio Martinoli: *SwisTrack: a tracking tool for multi-unit robotic and biological systems*. In: *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vols 1-12*, Seite 2185–2191, 2006. IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, PEOPLES R CHINA, OCT 09-13, 2006. (Zitiert auf Seite 223)
- [627] Boltes, Maik et al. (Herausgeber): *Workshop on Collecting and Analysing Experimental Data of Pedestrian Dynamics*, 2014. <http://ped.fz-juelich.de/workshop>. (Zitiert auf den Seiten 223 und 292)
- [628] Waldau, Nathalie, Peter Gattermann, Hermann Knoflacher und Michael Schreckenberg (Hrsg.): *Pedestrian and Evacuation Dynamics 2005*. Springer Berlin Heidelberg, 2007. (Zitiert auf den Seiten 226, 249 und 265)
- [629] *Interflam 2007: Proceedings of the Eleventh International Conference*, Nummer Bd. 1 in *Interflam 2007: Proceedings of the Eleventh International Conference*. Interscience Communications, 2007. (Zitiert auf den Seiten 226, 266 und 274)
- [630] *13th IEEE International Conference on Computer Vision (ICCV)*, Barcelona, Spain, 2011. (Zitiert auf den Seiten 227, 228, 232, 240, 242, 244, 248, 250, 253 und 259)

- [631] Weidmann, Ulrich, Uwe Kirsch und Michael Schreckenberg (Hrsg.): *Pedestrian and Evacuation Dynamics 2012*. Springer Berlin Heidelberg, 2014. (Zitiert auf den Seiten 227 und 242)
- [632] Peacock, Richard D., Erica D. Kuligowski und Jason D. Averill (Hrsg.): *Pedestrian and Evacuation Dynamics 2010*. Springer Berlin Heidelberg, 2011. (Zitiert auf den Seiten 235, 266, 267, 268, 273 und 275)
- [633] Klingsch, Wolfram W. F., Christian Rogsch, Andreas Schadschneider und Michael Schreckenberg (Hrsg.): *Pedestrian and Evacuation Dynamics 2008*. Springer Berlin Heidelberg, 2010. (Zitiert auf den Seiten 237, 260, 267, 269, 273 und 274)
- [634] Hoogendoorn, Serge, Winnie Daamen und Dorine Duives (Hrsg.): *Pedestrian and Evacuation Dynamics 2014*. Springer Berlin Heidelberg, 2014. (Zitiert auf den Seiten 269 und 275)
- [635] Batchelor, Bruce G.: *Coming to terms with machine vision and computer vision: they're not the same!* Advanced Imaging, Seite 22–26, 1999. (Zitiert auf Seite 277)
- [636] Boltes, Maik und Armin Seyfried: *Traffic and Granular Flow Conference celebrates 10th Edition by returning to Jülich*. Innovatives Supercomputing in Deutschland (inSiDE), 11(2):110, 2013. (Zitiert auf Seite 281)
- [637] Heesch, Dimitri van: *Doxygen*. <http://www.stack.nl/~dimitri/doxygen/index.html> (15.02.2014). (Zitiert auf Seite 290)
- [638] Research, Point Grey: *censys3D*. <http://ptgrey.com/products/censys3d/samples.asp> (25.11.2013). (Zitiert auf Seite 299)
- [639] Research, Point Grey: *people tracking - censys3D*. Technical Report, Point Grey Research, 2000. (Zitiert auf Seite 299)



Teilveröffentlichungen

Liste eigener Veröffentlichungen, die während der Arbeit an der Dissertation entstanden und zum Thema der Fußgängerdynamik gehören, beginnend mit dem Autor dieser Arbeit als Erstautor (ergänzt um den jeweilig neuen Aspekt in den aufeinander aufbauenden Arbeiten) und sortiert nach Erscheinungsjahr:

- Ermittlung von Laufwegen markierter Personen in der Ebene:
Boltes, Maik, Armin Seyfried, Bernhard Steffen und Andreas Schadschneider: *Automatic Extraction of Pedestrian Trajectories from Video Recordings*. In: KLINGSCH, W. W. F. et al. [633], Seite 43–54, 2010.
- Zusammenspiel von Laufwegeerfassung, Laufwegeanalyse und Modellbildung:
Boltes, Maik, Mohcine Chraïbi, Ulrich Kemloh und Jack Liddle: *Tracking, Analysis and Simulation of Crowds*. ASIM-Workshop - Trends in Computational Science and Engineering: Foundations of Modeling and Simulation, 2010. (Best Poster Award)
- Ermittlung von Laufwegen markierter Personen an Treppen durch Stereoaufnahmen:
Boltes, Maik, Armin Seyfried, Bernhard Steffen und Andreas Schadschneider: *Using Stereo Recordings to Extract Pedestrian Trajectories Automatically in Space*. In: PEACOCK, RICHARD D. et al. [632], Seite 751–754, 2011.
- Verbesserung von Laufwegen markierter Personen durch Berücksichtigung ihres Höhenfeldes:
Boltes, Maik, Jun Zhang, Armin Seyfried und Bernhard Steffen: *T-junction: Experiments, trajectory collection, and analysis*. In: *IEEE International Conference on Computer Vision Workshop on Modeling, Simulation and Visual Analysis of Large Crowds*, Seite 158–165, 2011.
- Zusammenfassung der Methoden zur Laufwegeerfassung:
Boltes, Maik und Armin Seyfried: *Collecting Pedestrian Trajectories*. Neurocomputing, Special Issue on Behaviours in Video, 100:127–133, 2013.

- Bericht über die Durchführung der TGF Konferenz 2013 in Jülich :
Boltes, Maik und Armin Seyfried: *Traffic and Granular Flow Conference celebrates 10th Edition by returning to Jülich*. Innovatives Supercomputing in Deutschland, 11(2):110, 2013.
- Zusammenfassung der Methoden und Ergebnisse zur Laufwegeerfassung und -analyse:
Boltes, Maik, Jun Zhang und Armin Seyfried: *Analysis of crowd dynamics with laboratory experiments*. In: ALI, SAAD et al. [112], Kapitel 4, Seite 67 – 98. ISBN: 978-1-4614-8482-0 (Print) 978-1-4614-8483-7 (Online), 2013.
- Ermittlung von Laufwegen in Stereoaufnahmen zur Ermöglichung von Feldstudien:
Boltes, Maik und Armin Seyfried: *Tracking People in Crowded Scenes*. In: *Pedestrian and Evacuation Dynamics 2012*, 533–542, 2014.
- Seyfried, Armin, Bernhard Steffen, Wolfram Klingsch und Maik **Boltes**: *The fundamental diagram of pedestrian movement revisited*. Journal of Statistical Mechanics: Theory and Experiment, P10002, 2005.
- Seyfried, Armin, Bernhard Steffen, Wolfram Klingsch, Thomas Lippert und Maik **Boltes**: *Steps Toward the Fundamental Diagram - Empirical Results and Modelling*. In: Waldau, N., P. Gattermann, H. Knoflacher und M. Schreckenberg (Hrsg.): *Pedestrian and Evacuation Dynamics 2005*, Seite 377–390, 2007.
- Seyfried, Armin, Bernhard Steffen, Wolfram Klingsch, Thomas Lippert und Maik **Boltes**: *The Fundamental Diagram of Pedestrian Movement Revisited - Empirical Results and Modelling*. In: Schadschneider, A., T Pöschel, R. Kühne, M. Schreckenberg und D. E. Wolf (Hrsg.): *Traffic and Granular Flow 2005*, Seite 305–314, 739, 2007.
- Rupperecht, Tobias, Armin Seyfried, Wolfram Klingsch und Maik **Boltes**: *Bottleneck capacity estimation for pedestrian traffic*. In: *Interflam 2007* [629], Seite 1423–1430, 2007.
- Seyfried, Armin, Tobias Rupperecht, Andreas Winkens, Oliver Passon, Bernhard Steffen, Wolfram Klingsch und Maik **Boltes**: *Capacity Estimation for emergency exits and bottleneck*. In: *Interflam 2007* [629], Seite 247–258, 2007.
- Seyfried, Armin, Bernhard Steffen, Andreas Winkens, Tobias Rupperecht, Maik **Boltes** und Wolfram Klingsch: *Empirical data for pedestrian flow through bottlenecks*. In: Appert-Rolland, Cécile, Francois Chevoir, Philippe Gondret, Sylvain Lassarre, Jean-Patrick Lebacque und Michael Schreckenberg (Hrsg.): *Traffic and Granular Flow '07*, Seite 189–199, 2009.
- Seyfried, Armin, Oliver Passon, Bernhard Steffen, Maik **Boltes**, Tobias Rupperecht und Wolfram Klingsch: *New insights into pedestrian flow through bottlenecks*. Transportation Science, 43(3):395–406, 2009.
- Seyfried, Armin, Maik **Boltes**, Jens Kähler, Wolfram Klingsch, Andrea Portz, Tobias Rupperecht, Andreas Schadschneider, Bernhard Steffen und Andreas Winkens: *Enhanced empirical data for the fundamental diagram and the flow through bottlenecks*. In: KLINGSCH, W. W. F. et al. [633], Seite 145–156, 2010.

- Klüpfel, Hubert, Armin Seyfried, Stefan Holl, Maik **Boltes**, Mohcine Chraibi, Ulrich Kemloh, Andrea Portz, Jack Liddle, Tobias Rupprecht, Andreas Winkens, Wolfram Klingsch, Christian Eilhardt, Stefan Nowak, Andreas Schadschneider, Tobias Kretz und Michael Krabbe: *HERMES - Evacuation Assistant for Arenas*. In: *Future Security*, 2010.
- Holl, Stefan, Armin Seyfried und Maik **Boltes**: *Hermes - Erforschung eines Evakuierungsassistenten für den Krisenfall bei Großveranstaltungen*. In: Fähnrich, K. P. und B. Franczyk (Hrsg.): *IT-Unterstützung von Rettungskräften, Informatik 2010, Service Science - Neue Perspektiven für die Informatik*, Band 2, Seite 162–167, 2010.
- Steffen, Bernhard, Armin Seyfried und Maik **Boltes**: *Reliability issues in the microscopic modeling of pedestrian movement*. In: *QMath 11 - Mathematical Results in Quantum Physics*, WSPC, 2010.
- Liddle, Jack, Armin Seyfried, Bernhard Steffen, Wolfram Klingsch, Tobias Rupprecht, Andreas Winkens und Maik **Boltes**: *Microscopic insights into pedestrian motion through a bottleneck, resolving spatial and temporal variations*. 2011.
- Steffen, Bernhard, Maik **Boltes** und Armin Seyfried: *Improved Methods for Checking Forces Based Models of Pedestrian Dynamics*. In: PEACOCK, RICHARD D. et al. [632], Seite 885–888, 2011.
- Steffen, Bernhard, Armin Seyfried und Maik **Boltes**: *Reliability issues in the microscopic modeling of pedestrian movement*. In: Pavel Exner, Hradec Kralove (Herausgeber): *Mathematical results in Quantum Physics*, Seite 254–259, 2011.
- Zhang, Jun, Wolfgang Mehner, Erik Andresen, Stefan Holl, Maik **Boltes**, Andreas Schadschneider und Armin Seyfried: *Comparative analysis of pedestrian, bicycle and car traffic moving in circuits*. In: *2nd Conference of Transportation Research Group of India (2nd CTRG)*, 2013.
- Zhang, Jun, Wolfgang Mehner, Stefan Holl, Maik **Boltes**, Erik Andresen, Andreas Schadschneider und Armin Seyfried: *Universal flow-density relation of single-file bicycle, pedestrian and car motion*. *Physics Letters A*, 2014.
- Liao, Weichen, Armin Seyfried, Jun Zhang, Maik **Boltes**, Xiaoping Zheng und Ying Zhao: *Experimental Study on Pedestrian Flow through Wide Bottleneck*. In: HOOGENDOORN, SERGE et al. [634], 2014.
- Seyfried, Armin, Erik Andresen, Maik **Boltes**, Stefan Holl, Wolfgang Mehner, Andreas Schadschneider und Jun Zhang: *Universalities in fundamental diagrams of cars, bicycles and pedestrians*. In: *Summer Symposium Celebrating 50 Years of Traffic Flow Theory*, 2014.
- **Boltes**, Maik , Mohcine Chraibi, Stefan Holl, Armin Ulrich Kemloh Wagoum, Gregor Lämmel, Weichen Liao, Wolfgang Mehner, Antoine Tordeux und Jun Zhang: *Experimentation, data collection, modeling and simulation of pedestrian dynamics*. In: *Statistics, probability and numerical analysis*, 2014.
- Chraibi, Mohcine, Maik **Boltes**, Andreas Schadschneider und Armin Seyfried (Hrsg.): *Traffic and Granular Flow 2013*. Springer, 2015.



Zeitschriften und Konferenzen

Da das Thema dieser Arbeit ein Anwendungsgebiet der Computer Vision (CV) im Bereich der Fußgängerdynamik ist, sind wissenschaftliche Veröffentlichungen in diesem Gebiet in beiden Feldern und verbindenden Forschungsgebieten (z. B. Sicherheitstechnik, Robotik) zu finden. Für den englischen Begriff Computer Vision existiert keine in dem Forschungsgebiet verstandene deutsche Übersetzung. Computer Vision ist verwandt bzw. wird manchmal gleich gesetzt mit den Disziplinen Bildanalyse (*image analysis*), Bildverstehen oder maschinellem Sehen (*machine vision*, MV), wobei das maschinelle Sehen als praktische Umsetzung von Verfahren aus dem Bereich Computer Vision verstanden werden kann [635].

B.1 Zeitschriften und Bücher

Die zitierten Publikationen aus dem Bereich der *Computer Vision* entstammen hauptsächlich folgenden Zeitschriften:

- Computer Vision and Image Understanding
- Computer Vision Systems
- IEEE Transactions on Pattern Analysis and Machine Intelligence
- IEEE Computer Society Conference on Computer Vision and Pattern Recognition
- Image and Vision Computing
- International Journal of Computer Vision
- Pattern Recognition Letters
- Pattern Recognition
- Advances in Complex Systems

- Pattern Recognition and Artificial Intelligence
- Artificial Intelligence Research

Des Weiteren findet man anwendungsnahe Veröffentlichungen zur *Analyse von Verkehrssystemen* in:

- Transportation Science
- Fire Safety Journal
- Journal of Statistical Mechanics: Theory and Experiment
- Journal of Transportation Engineering
- Physica A: Statistical Mechanics and its Applications
- Transportation Research Part B: Methodological

Speziell dem Thema *Bildanalyse von Menschenmengen* widmen sich folgende Sonderausgaben der Zeitschriften:

- IAPR Pattern Recognition Letters,
Special Issue on Pattern Recognition and Crowd Analysis [114]
- Neurocomputing,
Special Issue on Behaviours in Video [111]
- Eurasip Journal in Image and Video Processing,
Special Issue on Anthropocentric Video Analysis: Tools and Applications [102]

Bücher zum Thema Personenerkennung:

- Pedestrian Behavior: Models, Data Collection and Applications [104]
- Visual Analysis of Humans – Looking at People [108]
- Modeling, Simulation and Visual Analysis of Crowds [112]

B.2 Konferenzen und Schulen

Die wichtigen Konferenzen, Kongresse, Symposien und Schulen im Bereich der Computer Vision und Fußgängerdynamik sind im folgenden aufgelistet. Die Vielzahl an Veranstaltungen spiegelt die schnelle Entwicklung in diesem Bereich wider.

Computer Vision Konferenzen:

- ICCV, IEEE International Conference on Computer Vision,
Sydney, Australien, 8.-15.12.2013

- **CVPR**, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Columbus, Vereinigte Staaten, 24.-27.06.2014
- **ECCV**, European Conference on Computer Vision, Zürich, Schweiz, 6.-12.09.2014
- **ICPR**, International Conference on Pattern Recognition, Stockholm, Schweden, 24.-28.08.2014
- **ACCV**, Asian Conference on Computer Vision, Singapur, Singapur, 1.-5.11.2014
- **VISAPP**, International Conference on Computer Vision Theory and Applications, Berlin, Deutschland, 11.-14.03.2015
- **ICIAR**, International Conference on Image Analysis and Recognition, Vilamoura, Portugal, 22.-24.10.2014
- **CIARP**, Iberoamerican Congress on Pattern Recognition, Puerto Vallarta, Mexiko, 2.-5.11.2014
- **IAPR**, Asian Conference on Pattern Recognition, Okinawa, Japan, 5.-8.11.2013
- **PSIVT**, Pacific-Rim Symposium on Image and Video Technology, Auckland, Neuseeland, 23.-27.11.2015
- **ICIP**, IEEE International Conference on Image Processing, Melbourne, Australien, 15.-18.09.2013
- **CAIP**, Computer Analysis of Images and Patterns, York, Großbritannien, 27.-29.08.2013
- **ICVS**, International Conference on Computer Vision Systems, Kopenhagen, Dänemark, 6.-9.07.2015
- **IPCV**, International Conference on Image Processing, Computer Vision, and Pattern Recognition, Las Vegas, Vereinigte Staaten, 21.-24.07.2014
- **ICISP**, International Conference on Image and Signal Processing, Bangalore, Indien, 9.-11.08.2013
- **BMVC**, British Machine Vision Conference, Nottingham, Großbritannien, 1.-5.09.2014
- **CGIV**, Computer Graphics, Imaging and Visualization, Macau, China, 6.-8.08.2013
- **CGI**, Computer Graphics International, Sydney, Australien, 10.-13.06.2014

- **ISVC**, International Symposium on Visual Computing, Rethymnon, Kreta, Griechenland, 29.-31.07.2013
- **MIRAGE**, International Conference on Computer Vision / Computer Graphics Collaboration Techniques and Applications, Berlin, Deutschland, 6.-7.06.2013
- **SIBGRAPI**, Conference on Graphics, Patterns, and Images, Arequipa, Peru, 5.-8.08.2013
- **ICRA**, IEEE International Conference on Robotics & Automation, Karlsruhe, Deutschland, 6.-10.05.2013

Computer Vision Schulen, Workshops und Messen:

- **ICVSS**, International Computer Vision Summer School, Catania, Italien, 14.-20.07.2013
- **WSCG**, International Conference on Computer Graphics, Visualization and Computer Vision (zuvor: Winter School of Computer Graphics), Pilsen, Tschechien, 8.-12.06.2015
- **WACV**, IEEE Workshop on Application of Computer Vision, Clearwater, Vereinigte Staaten, 17.-18.01.2013
- **VMV**, International Workshop on Vision, Modeling and Visualization, Darmstadt, Deutschland, 8.-10.10.2014
- **IWAIT**, International Workshop on Advanced Image Technology, Nagoya, Japan, 7.-9.01.2013
- **VISION**, Leading world trade fair for machine vision, Stuttgart, Deutschland, 4.-6.11.2014
- **AUTOMATICA**, International Trade Fair for Automation and Mechatronics, München, Deutschland, 20.-23.05.2014

Computer Vision mit der Anwendung auf Menschenmengen:

- Workshop on Modeling, Simulation and Visual Analysis of Large Crowds, Barcelona, Spanien, 7.11.2013
- **PRCA**, Workshop on Pattern Recognition and Crowd Analysis, Tsukuba, Japan, 11.11.2012

Fußgängerdynamik:

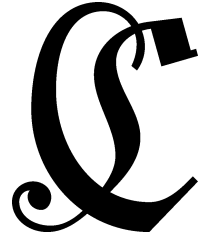
- **PED**, International Conference on Pedestrian and Evacuation Dynamics, Delft, Niederlande, 22.-24.10.2014

- **TGF**, Traffic and Granular Flow, Jülich, Deutschland, 25.-27.09.2013 [636]
- **TRB**, TRB Annual Meeting, Washington, Vereinigte Staaten, 12.-16.01.2014
- **NATEC**, North American Travel Monitoring Exposition and Conference: Improving Traffic Data Collection, Analysis, and Use Chicago, Vereinigte Staaten, 29.6.-2.07.2014
- **ITSC**, International IEEE Conference on Intelligent Transportation Systems, Qingdao, China, 8-11.10.2014
- **CTRG**, Conference of Transportation Research Group of India, Kolkata, Indien, 17.-20.12.2015
- **CASA**, Conference on Computer Animation and Social Agents, Houston, Vereinigte Staaten, 26.-28.05.2014
- **ACRI**, Cellular Automata for Research and Industry, Krakau, Polen, 22.-25.09.2014
- **AAMAS**, International Conference on Autonomous Agents and Multiagent Systems, Paris, Frankreich, 05.-06.05.2014
- Symposium Celebrating 50 Years of Traffic Flow Theory, Portland, Oregon, Vereinigte Staaten, 11.-13.08.2014
- **ICNAAM**, International Conference of Numerical Analysis and Applied Mathematics, Rhodos, Griechenland, 22.-28.09.2014
- Workshop on pedestrian models, Lausanne, Schweiz, 10.-12.04.2014
- International Scientific Conference of Young Scientists and Specialists „Problems of Technosphere Safety“, Moskau, Russland, 08.04.2014
- **ICEM**, International Conference on Evacuation Modeling and Management, Tainan, Taiwan, 01.-03.06.2015
- **IAFSS**, International Association for Fire Safety Science Symposium, Lund, Schweden, 06.2017
- **CUEE**, International Conference on Civil, Urban and Environmental Engineering, Beijing, China, 19.-20.08.2014
- **CCS**, Workshop on Complex Collective Systems, Krakau, Polen, 6.-9.09.2015
- Interaction Networks and Collective Motion in Swarms, Flocks and Crowds, Espoo, Finnland, 26.-28.05.2014

- Human Behaviour in Fire Symposium,
Cambridge, Großbritannien, 28.-30.09.2015
- Walk21 Vienna stepping ahead Conference,
Wien, Österreich, 20.-23.10.2015

Überwachungs- und Sicherheitstechnik:

- BMBF-Innovationsforum „Zivile Sicherheit“,
Berlin, Deutschland, 7.-9.05.2014
- Kongress des Verbandes für Sicherheitstechnik,
Leipzig, Deutschland, 8.-9.04.2014
- AVSS, IEEE International Conference On Advanced Video and Signal Based Surveillance,
Seoul, Korea, 26.-29.08.2014
- IT-Unterstützung im Emergency Management & Response,
Koblenz, Deutschland, 17.09.2013
- Future Security Conference,
Berlin, Deutschland, 16.-18.09.2014
- **Showtech**, Internationale Fachmesse und Kongress für Theater, Film und Event,
Berlin, Deutschland, 19.-21.05.2015



Videos

Da diese Arbeit im Kern die Analyse von Bildsequenzen beschreibt, werden in diesem Abschnitt einige zu analysierende Videos und Videos zur Visualisierung der Resultate aufgelistet. Die Liste ist alphabetisch. Eine Zuordnung zu den Versuchsreihen kann Tab. 4.3 entnommen werden und ist zudem hinter den Elementen der Aufzählung vermerkt. Folgende Videos sind keiner Versuchsreihe zuzuordnen, da

- sie Ergebnisse von mithilfe der Experimente entwickelte Simulationen zeigen:
33, 46, 51, 52, 53,
- eine Komposition aus verschiedenen Inhalten darstellen:
49,
- sie Menschenmengen zeigen, die von anderen aufgezeichnet wurden:
37, 40.

Simulationen sind zudem in Videos zu sehen, bei denen Experimente mit Simulationsergebnissen verglichen werden: 16, 17, 18, 24, 43.

- [1] Geringe Breitenvariation ändert den Personenfluss (Marker Nr. 1)¹ (Exp. Nr. 3)²:
090-100-110.wmv³
- [2] Linienbildung verschwimmt bei zunehmender Korridorbreite (Exp. Nr. 3):
090-160-250.wmv
- [3] Engstellenexperiment mit einer kreisförmigen Anfangskonfiguration (Exp. Nr. 6):
ao-360-400.wmv

¹Nummern beziehen sich auf die Markernummer auf Seite 57 bzw. Abb. 3.8 und 3.9

²Experimentnummern beziehen sich auf Versuchsreihen aus Tab. 4.1

³In der PDF-Version des Dokumentes kann zum Starten des Videos dieses angeklickt werden, ansonsten muss folgende URL jeder Datei vorangestellt werden:
<http://ped.fz-juelich.de/boltes/diss/videos/>

- [4] Methoden zur Dichtebestimmung: klassisch und Voronoi (Exp. Nr. 3):
b090-DichteKlassischVoronoi.wmv
- [5] Experiment zur Räumung eines mit Absperrungen abgetrennten Areals (Marker Nr. 2) (Exp. Nr. 9):
barrier-hoheDichte.wmv
- [6] Bidirektionaler Personenverkehr in einem Korridor der Breite 4 m (Exp. Nr. 9):
bi-corr-400-hoheDichte.wmv
- [7] Bidirektionaler Personenverkehr in einem Korridor der Breite 3,6 m (Exp. Nr. 6):
bo-360-200-200.wmv
- [8] Variation der Engstellenbreite (Exp. Nr. 2):
breitenvariationRotunde.wmv
- [9] Kreuzung mit großem Personenaufkommen (Exp. Nr. 9):
crossing-90-hoheDichte.wmv
- [10] Kreuzung mit großem Personenaufkommen, die aufgrund einer Verstopfung geräumt wird; Aufnahme mit Fischaugenkamera (Exp. Nr. 9):
crossing-90-hoheDichte-stopp.wmv
- [11] Kreuzung mit großem Personenaufkommen und der visuellen Anweisung und baulichen Unterstützung zur Nutzung eines Kreisverkehrs (Exp. Nr. 9):
crossing-90-saeule-hoheDichte.wmv
- [12] Dreierkreuzung mit großem Personenaufkommen (Exp. Nr. 9):
crossing-120-hoheDichte.wmv
- [13] Dreierkreuzung mit großem Personenaufkommen und der visuellen Anweisung und baulichen Unterstützung zur Nutzung eines Kreisverkehrs (Exp. Nr. 9):
crossing-120-hoheDichte-kreisverkehr.wmv
- [14] Fortlaufendes Differenzbild aufeinanderfolgender Einzelbilder (Exp. Nr. 2):
differenzbild.wmv
- [15] Markerlose Erkennung einer virtuellen Person:
einePersonVirtuell.wmv
- [16] Vergleich von Experiment und kalibrierter Simulation auf Basis des „Generalized centrifugal force model“ (GCFM) [55] (Exp. Nr. 3):
engstelleVergleichExpSim.wmv
- [17] Vergleich von Experiment und Simulation durch Bildüberlagerung (Exp. Nr. 3):
engstelleVergleichExpSimUeberlagerung.wmv
- [18] Vergleich von Experiment und Simulation mit Varianten zur Modellierung des Volumenaus-schlusses (GCF-Modell) (Exp. Nr. 3):
engstelleVergleichExpSimVarianten.wmv
- [19] Dichtevariation im eindimensionalen Fahrradverkehr (Exp. Nr. 7):
fahrradDichtevariation.wmv

-
- [20] Erkennung Fahrradfahrer mittels Farbmarker (Marker Nr. 4) (Exp. Nr. 7):
fahrradErkennungOben.avi
- [21] Erkennung Fahrradfahrer mittels individuellen Marker (Marker Nr. 5) (Exp. Nr. 7):
fahrradErkennungSeite.avi
- [22] Aufnahme von Mundloch in ESPRIT Arena bei Feldstudie:
feldstudieArena.wmv
- [23] Dichtevariation im eindimensionalen Fußgängerverkehr (Marker Nr. 1) (Exp. Nr. 3):
fussgaengerDichtevariation.wmv
- [24] Dichtevariation im eindimensionalen Fußgängerverkehr: real und Simulation (Exp. Nr. 3):
fussgaengerDichtevariationMitSimulation.wmv
- [25] Vier Arten von Versuchen in künstlichen Aufbauten (Exp. Nr. 6):
hermesMesseVersucheUeberblick.wmv
- [26] Einfluss der Kultur auf den Personenfluss (Exp. Nr. 1 und 4):
indienDeutschland.wmv
- [27] Gangverhalten in einer Personenschlange (Marker Nr. 4) (Exp. Nr. 8):
japanLinie14.wmv
- [28] Raumein- und austritt (Marker Nr. 3) (Exp. Nr. 8):
japanRaum10.wmv
- [29] Kanten hervorhebung bei Engstellenversuch (Exp. Nr. 2):
kantendetektion.wmv
- [30] T-Kreuzung mit visualisierten getrackten Personen (Marker Nr. 2) (Exp. Nr. 6):
ko-240-050-240-tracking.wmv
- [31] Dichte- und Geschwindigkeitsermittlung mittels Voronoi-Methode an einer T-Kreuzung
(Exp. Nr. 6):
ko-240-240-240-dichteGeschwindigkeit.wmv
- [32] T-Kreuzung mit großem Personenaufkommen (Exp. Nr. 6):
ko-300-150-300.wmv
- [33] Simulation einer T-Kreuzung mit dem GCF-Modell (Exp. Nr. 6):
ko-350-sim.avi
- [34] Personenerkennung ohne Marker (Exp. Nr. 6):
ko-ohneMarker.wmv
- [35] 3D-Visualisierung mit Hilfe der Software `travisto` von experimentell ermittelten Laufwegen
(Exp. Nr. 6):
ko-travistoVisualisierung.wmv
- [36] Längenvariation des Korridors nach einer Engstelle (Exp. Nr. 3):
laengenvariation.wmv

- [37] Turbulenzen der Menschenmenge bei der Loveparade kurz vor dem Unglück:
loveparade.wmv
- [38] Zoom auf ein Experiment zur Veranschaulichung der Aufnahmen von oben (Exp. Nr. 2):
menschenVonOben.wmv
- [39] Verlassen der Tribüne im Oberrang der ESPRIT Arena durch ein Mundloch (Exp. Nr. 6):
moll-arenaUeberblick.wmv
- [40] Orte von großen Fußgängerströmen:
motivation.wmv
- [41] Steigerung der Motivation erhöht den Personenfluss (Exp. Nr. 5):
motivationWuppertal.wmv
- [42] Entleerung einer Tribüne im Unterrang in der ESPRIT Arena über ein Mundloch (Exp. Nr. 6):
mull.wmv
- [43] Vergleich von Experiment und Simulation beim Verlassen einer Tribüne (Exp. Nr. 6):
mull-experimentTrajektorienSimulation.wmv
- [44] Ablauf der Erfassung der Laufwege jeder Person während des Verlassens einer Tribüne über ein Mundloch (Marker Nr. 2) (Exp. Nr. 6):
mull-petrackAblauf.wmv
- [45] Verlassen der Tribüne im Unterrang der ESPRIT Arena durch ein Mundloch und die erfassten Laufwege (Marker Nr. 2) (Exp. Nr. 6):
mull-realPetrack.wmv
- [46] Simulation mit dem GCF-Modell des Verlassens einer Tribüne (Exp. Nr. 6):
mull-simulation.wmv
- [47] Laufwege von einem Experiment der Tribünenentleerung; die Farbe codiert die Geschwindigkeit (Exp. Nr. 6):
mull-trajektorienGeschwindigkeitfarbe.wmv
- [48] Laufwege von einem Experiment der Tribünenentleerung; die Farbe entspricht der Shirtfarbe der Probanden zur Separation der Personenströme (Exp. Nr. 6):
mull-trajektorienShirtfarbe.wmv
- [49] Zusammenschnitt von Videos zur Veranschaulichung des Prozesses vom Experiment über die Analyse bis zum Modell :
ped.wmv
- [50] Ablauf der Erfassung der Laufwege: Entzerrung, Vermessung, Erkennung, Verfolgung und Größenbestimmung (Marker Nr. 1) (Exp. Nr. 3):
petrack-ablauf.wmv
- [51] Simulation einer Evakuierung von 900 Personen aus einer vereinfachten Geometrie der ESPRIT Arena:
simArena-900personen.avi

- [52] Simulation der Evakuierung der Promenade der ESPRIT Arena:
`simArena-promenade-2500personen.avi`
- [53] Simulation der Evakuierung eines Tribünenblockes der ESPRIT Arena:
`simArena-raum80.wmv`
- [54] Dichte- und Geschwindigkeitsvergleich von uni- und bidirektionalen Personenströmen in einem Korridor (Exp. Nr. 6):
`uo-bo-dichteGeschwindigkeit.wmv`
- [55] Entwicklung der Voronoi-Dichte und dem integrierten Profil (Exp. Nr. 6):
`voronoiErklaerung.wmv`



Implementationsbeschreibung

Dieses Kapitel beschreibt die entwickelte Software `PeTrack` und das Hilfsprogramm `combine`, in die alle in dieser Arbeit beschriebenen selbst entwickelten Verfahren eingeflossen sind.

Zudem sind Hilfsmittel aufgeführt, die zur Erstellung dieser Arbeit herangezogen wurden.

D.1 Voraussetzungen

D.1.1 Hilfsmittel

Diese Arbeit wurde in \LaTeX gesetzt. Zur Erstellung des Literaturverzeichnisses wurde \BIBTeX genutzt, dessen Datenbank mit `JabRef` verwaltet wurde.

Die Datenauswertung wurde mit Python und den zugehörigen Bibliotheken `Numpy` und `SciPy` durchgeführt. Die Herkunft von Bildern aus fremden Publikationen ist immer unter dem Bild angegeben. Eigene Illustrationen sind mit `Adobe Illustrator` gezeichnet worden. Funktionen und Kurven wurden mit `Maple` und Diagramme mit `gnuplot`, `matplotlib` (Python-Bibliothek), `Visualization Toolkit (VTK)` und `Microsoft Excel` erstellt.

Die virtuellen Menschenmengen wurden mit `Panda3D` realisiert. `Adobe After Effects` wurde zu Beginn bei Versuchsreihe Nr. 2 zur Verfolgung manuell markierter Personen herangezogen. Die meisten der in Anhang C aufgeführten Videos wurden mit der Software `Adobe Premiere` geschnitten. Zur Dokumentation und zur Erstellung von Klassenhierarchien sowie Aufrufgraphen wurde `Doxygen` verwendet.

D.1.2 Programmbibliotheken

Neben diversen Fotos entspringen die meisten Bilder dieser Arbeit der eigenen Software `PeTrack`, die wiederum bei der Bildmanipulation auf den Programmbibliotheken `OpenCV`, `Qt` bzw. `Qwt` und `OpenGL` zurückgreift.

Zur Berechnung der Disparitäten kamen Verfahren der OpenCV-, der libelas- und der Triclops-Bibliothek des Kameraherstellers PointGrey zum Einsatz.

Die grafische Oberfläche wurde mit Qt und Qwt realisiert. Die nicht selbst implementierten Methoden der Bildverarbeitung und -analyse entstammen der OpenCV-Bibliothek.

D.2 Aufbau von PeTrack



Abb. D.1: Logo der Software PeTrack

Der Name des Programms **PeTrack** steht für „**P**edestrian **T**racking“. Die Software ist plattform-unabhängig in der Programmiersprache C++ programmiert. Lediglich einige Video-Codecs und Bibliotheken der Stereokamera-Hersteller sind auf das Betriebssystem Windows beschränkt. Im Einsatz ist **PeTrack** aktuell unter Linux und Windows. Die Software ist frei herunterzuladen und nutzbar [456]. Sie unterstützt als einzige Sprache Englisch und liegt aktuell in der Version 0.7 vor. Das Logo der Software ist in Abb. D.1 abgebildet.

Das Programm **PeTrack** besteht aktuell aus ca. 28400 Programmzeilen und beinhaltet 80 Klassen. Wesentliche Klassen sind:

Petrack: zentrale Klasse, die den Programmablauf steuert (siehe Abb. D.2),

Control, **Player**, ***Item**, **View***: grafische Benutzeroberfläche (siehe Abb. D.3); Repräsentation von Kontrollstrukturen und Bildelementen,

Animation, ***AviFile**: zur Verwaltung der Bildsequenzen,

***Filter**: Filter, die vor dem Erkennungsschritt auf das ganze Bild angewendet werden können, z. B. zur Steuerung der Helligkeit, des Kontrastes, des Weißabgleichs, der Entzerrung, der Rotation, der Spiegelung und der Rahmenbreite,

MarkerCasern*, **MarkerHermes***, **MarkerJapan***, **MarkerColor***, **Person***: Gruppen von Klassen zur Erkennung, Steuerung und Darstellung der entsprechenden Marker; die Klassen **Person*** dienen der markerlosen Erkennung,

Track*: zur Steuerung, Optimierung, Berechnung, Umwandlung und Darstellung der Laufwege,

Stereo*: zur Verwaltung, Steuerung, Auswertung und Darstellung von Stereoaufnahmen.

Abb. D.2 zeigt einen Aufrufgraphen des Konstruktors der Klasse **Petrack** und steht stellvertretend für die gesamte Softwaredokumentation, die mit Hilfe von Doxygen [637] erstellt wurde.

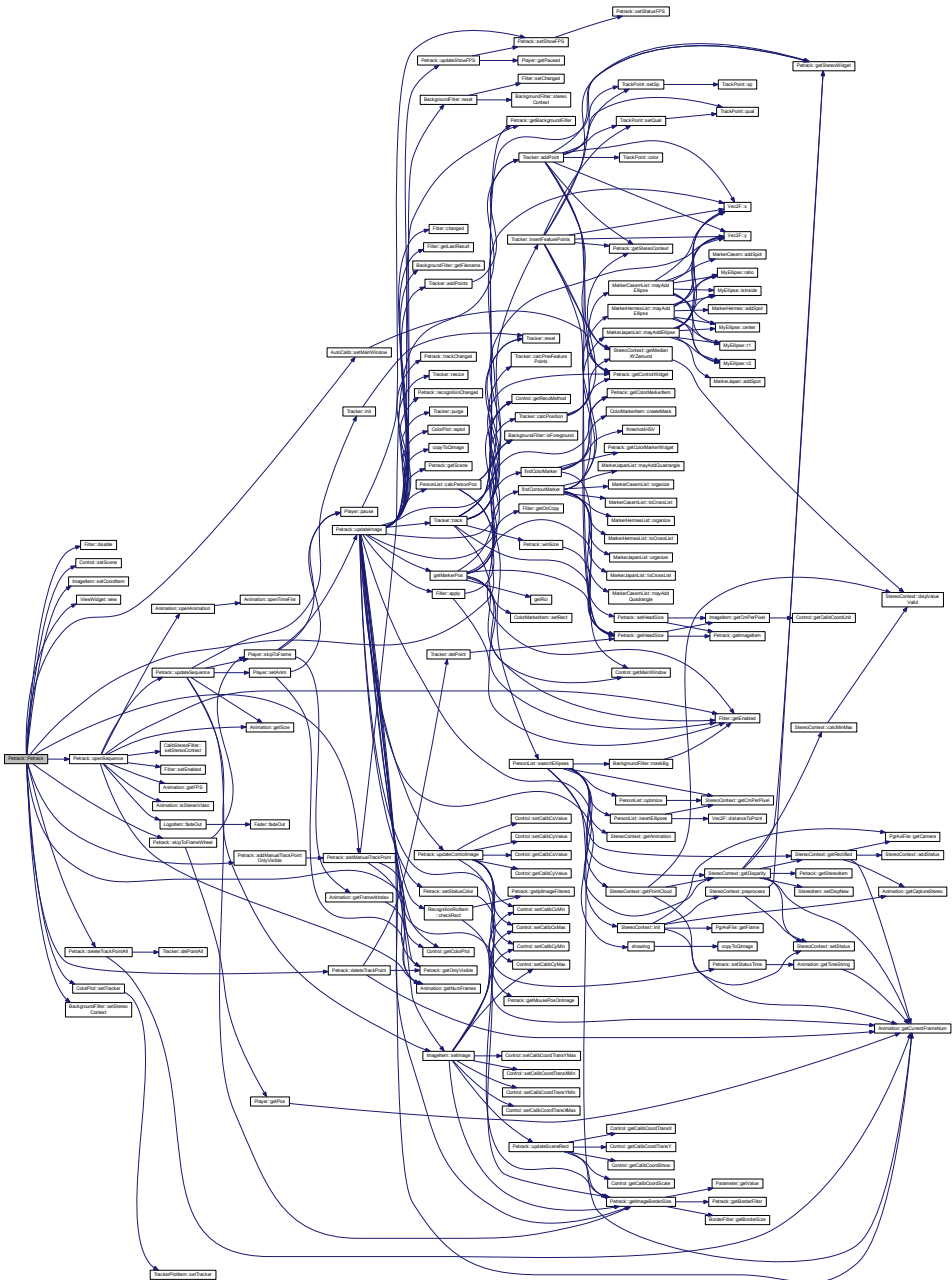


Abb. D.2: Aufrufgraph ausgehend vom Konstruktor der Klasse `PetTrack`

D.3 Nutzung von PeTrack

An dieser Stelle kann nur ein grober Überblick über die Bedienung der Software PeTrack gegeben werden (siehe [456] für eine englische Einführung). Um die Möglichkeiten ausführlicher kennenzulernen, ist am 27.10.2014 ein erster ganztägiger Workshop zur Sammlung und Analyse von experimentellen Daten der Fußgängerdynamik [627] geplant, in dem zur Datenerhebung die Software PeTrack detailliert vorgestellt wird. Die einzelnen Schritte der Extraktion sind auch Video Nr. 44 und 50 zu entnehmen. Neben der offiziellen Internetseite [456] können weitere Informationen bzw. Hilfen der Readme-Datei der Installation und den Schnellinfos der grafischen Oberfläche entnommen werden. Die Nutzung von der Kommandozeile aus wird auch in Abschnitt 4.2 erläutert. Alle Optionen werden bei Aufruf von PeTrack mit der Option `-help|-?` ausgegeben. Es sind viele Tastenkombinationen zum schnelleren Arbeiten implementiert, die ebenfalls [456] zu entnehmen sind.

Nach dem Laden einer Einzelbildsequenz oder eines Videos einer monokularen Kamera oder Stereokamera erscheint das erste Bild der Sequenz im Hauptfenster des linken Bereichs der grafischen Benutzeroberfläche von PeTrack (siehe Abb. D.3). Informationen zum Video werden in der Titelleiste angezeigt. Eine Navigation ist zeitlich unterhalb und örtlich (Vergrößerung bzw. Verkleinerung und Rotation) rechts neben dem Hauptfenster möglich. In der Statusleiste werden Informationen zum Pixel unterhalb der Maus angezeigt wie die reale Position im Raum (nur bei Stereoaufnahmen), die absolute Zeit der Aufnahme des sichtbaren Bildes, der Abspielgeschwindigkeit, die Höhe über Boden für eine wählbare Bodenentfernung, die Pixelkoordinate und -farbe. Im rechten Bereich des Programmfensters liegen vier den einzelnen Schritten der Auswertung entsprechende Registerkarten, die es erlauben, die Parameter der Schritte der Prozesskette (Kalibrierung, Erkennung, Verfolgung, Analyse) zu wählen. Eine Speicherung aller eingestellten Werte ist über eine XML-Projektdatei möglich.

Abb. D.5 zeigt die die Schritte der Prozesskette abbildenden Registerkarten. Unter der Registerkarte

- a) **Kalibrierung** (*calibration*) können als Vorverarbeitungsschritte Farbkorrekturen durchgeführt, ein farbiger Rahmen hinzugefügt, eine Hintergrundsubtraktion gesteuert (siehe Abschnitt 3.2 und 3.5.1) und das Bild horizontal oder vertikal gespiegelt werden; die intrinsischen Parameter können automatisch bestimmt oder manuell eingestellt werden (siehe Abschnitt 3.1.2), wobei einige Annahmen wie quadratische Pixel, die Lage der optischen Achse in der Bildmitte oder das Vorkommen einer tangentialen Verzerrung festgelegt werden können; die extrinsischen Parameter werden durch Positionierung eines Koordinatensystems auf dem Boden angegeben (siehe Abschnitt 3.1.1); für die visuelle Kontrolle der Güte der Entzerrung kann ein anpassbares kartesisches Gitter über das Bild gelegt werden,
- b) **Erkennung** (*recognition*) kann die Rate der Erkennung und der Bereich, in dem die Erkennung durchgeführt werden soll, eingestellt werden; welcher Marker (siehe Abb. 3.8 und 3.9) im Ausschnitt gesucht werden soll ist wählbar und einige Marker (Nr. 2 und 4) erlauben die Eingabe weiterer Parameter (siehe Abb. D.4 a und b); die Helligkeit des Markers gibt den dunkelsten akzeptierten Untergrund der Kopfbedeckung an; es ist möglich, auch die Marker heranzuziehen, bei denen die einzelnen Markerelemente nicht alle erkannt wurden; bei der Größenbestimmung über ein Farbmarkerelement kann des Weiteren die Verteilung der erkannten Farben visualisiert und einer Größe zugewiesen werden, wobei graue Farbmarkerelemente separat behandelt werden können; die Durchschnittspersonengröße wird für alle Marker ohne einer Möglichkeit der Größenangabe oder Nutzung von Stereoaufnahmen herangezogen,

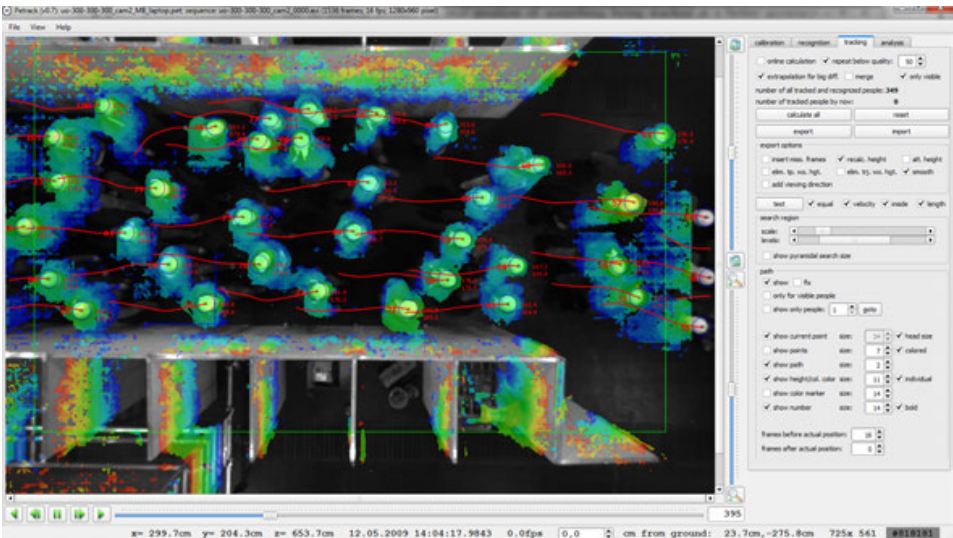


Abb. D.3: Grafische Benutzeroberfläche der Software PeTrack; sie gliedert sich in ein Hauptfenster mit der Bildsequenz und etwaigen Zusatzinformationen mit örtlichem Bezug, dem darunter liegenden Navigationsbereich und einer Statusleiste mit Angaben zum Pixel an der Mausposition, einer darüber liegenden Menüleiste (siehe Abb. D.4 c) und im rechten Bereich des Programmfensters Registerkarten der einzelnen Schritte der Auswertung (siehe Abb. D.5)

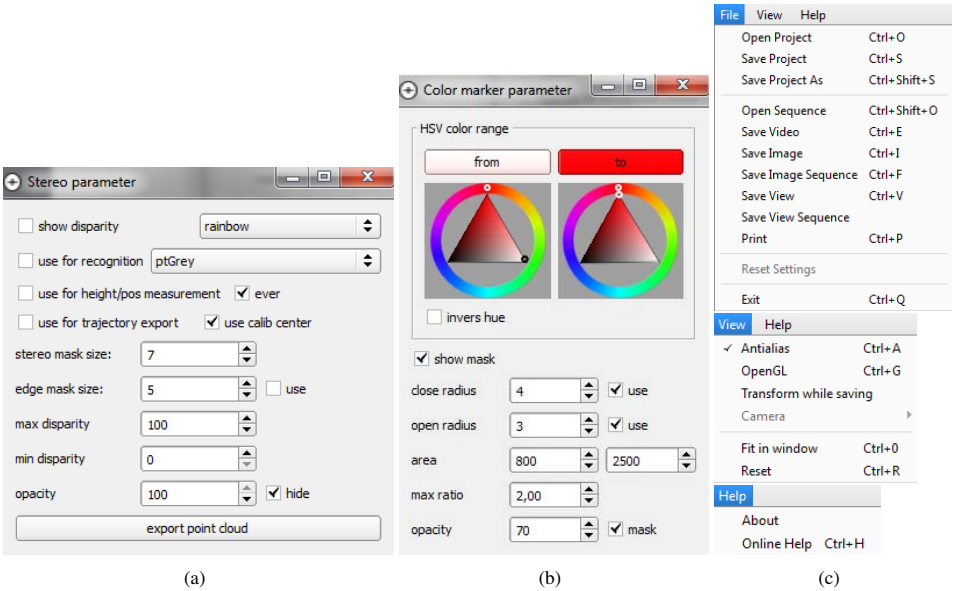


Abb. D.4: Eingabefenster für (a) die Steuerung der Berechnung der Disparitäten und (b) für die Parameterwahl des Farbmarkers; (c) Menüs der Menüleiste der grafischen Benutzeroberfläche von PeTrack

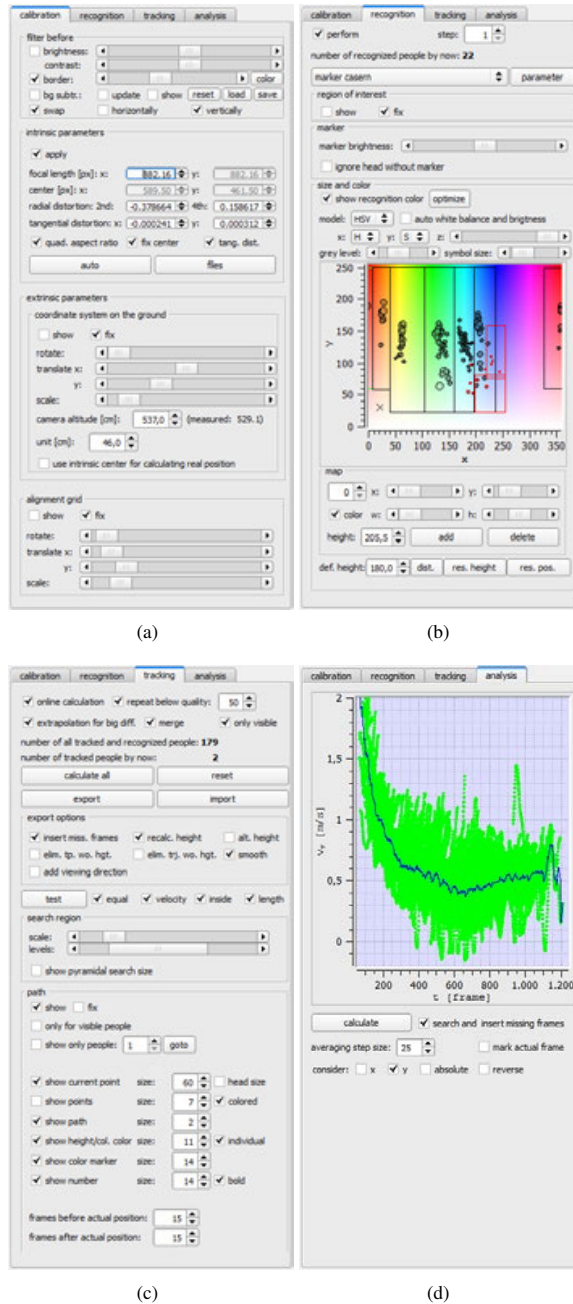


Abb. D.5: Registerkarten der grafischen Benutzeroberfläche von PetTrack, die die einzelnen Schritte der Auswertung der Bildsequenzen gliedern: (a) Kalibrierung, (b) Erkennung, (c) Verfolgung und (d) Analyse

- c) **Verfolgung** (*tracking*) kann eingestellt werden, ob und ab welcher Qualität für eine bereits verfolgte Person eine Verbesserung der Trajektorie durchgeführt werden soll (siehe Abschnitt 3.3.1); bei großen Änderungen in der Geschwindigkeit kann statt der getrackten Position eine extrapolierte Position gewählt werden (siehe Abschnitt 3.3.2); sollten sich zwei Laufwege sehr nahe kommen, so können diese verschmolzen werden; ist die Visualisierung der Trajektorien auf eine Auswahl beschränkt, so kann auch das Tracking auf diese Auswahl begrenzt werden; die automatische Berechnung aller Laufwege aller Personen sowie der Im- und Export dieser Trajektorien kann für verschiedene Datenformate angestoßen werden; beim Herausschreiben können Laufwege optimiert und korrigiert sowie auf Unregelmäßigkeiten hin getestet werden (siehe Abschnitt 3.3.4); die Größe und die Anzahl der Ebenen des pyramidenförmigen Suchbereichs des Lucas-Kanade-Merkmal-Verfolgers kann angepasst werden (siehe Abschnitt 3.3); alle wesentlichen Elemente der Laufwege können wahlweise visualisiert werden,
- d) **Analyse** (*analysis*) kann eine erste unmittelbare Analyse der Geschwindigkeit angeschlossen werden; die grünen Punkte zeigen die ggf. gemittelte Geschwindigkeit jeder Person, die blaue Linie ist die durchschnittliche Geschwindigkeit je Einzelbild; weitere Analysen werden für gewöhnlich außerhalb von PeTrack durchgeführt.

D.4 Nutzung von combine

PeTrack erlaubt die Personenverfolgung in einer Kameraperspektive. Um größere Bereiche abzudecken, bieten sich überlappende Perspektiven mehrerer Kameras an (siehe Abschnitt 4.1.1.2). Zur Verschmelzung der resultierenden Trajektorien aus mehreren Ansichten kann das separate Hilfsprogramm `combine` herangezogen werden. Dieses wird mit PeTrack ausgeliefert und ist nur über die Kommandozeile bedienbar. Die Verfahrensweise der Verschmelzung ist in Abschnitt 3.3.5 beschrieben. Dabei werden die Trajektorienpaare so gewählt, dass der Gesamtfehler des Quadratabstandes aller einander zugehörigen Punkte basierend auf [473] minimiert wird. Hierzu ist ein zeitlicher Versatz zwischen den Datensätzen zu bestimmen oder anzugeben. Zudem kann ein Datensatz transformiert werden, um den Fehler zu minimieren. Nach der Transformation kann eine automatische Neuzuordnung der Trajektorien veranlasst werden.

Der Aufruf kann mit folgenden Optionen und Argumenten erfolgen:

```
combine.exe [-help|-?] [-inBetween] [-autoAdjust]
            [-adjustWith x11 x12 x13 x21 x22 x23 x31 x32 x33 t1 t2 t3]
            [-alignWith x11 x12 x13 x21 x22 x23 x31 x32 x33 t1 t2 t3]
            [-addUnmatched] [-delay frames] [-varyingHeight]
            in1.txt in2.txt out.txt
```

`combine` fügt dabei `in1.txt` und `in2.txt` zusammen und erzeugt als Resultat `out.txt`. Die Optionen haben folgende Bedeutung:

- help|-?** : gibt eine Hilfe mit einer englischen Beschreibung zur Nutzung des Programms aus,
- inBetween** : veranlasst eine zweite Runde der Zuordnung der Trajektorien nach der Transformation eines Datensatzes,
- autoAdjust** : richtet den resultierenden Datensatz so aus, dass die Laufwege in ebenen Bereichen parallel zur Bewegungsebene verlaufen,

- adjustWith x_{11} x_{12} x_{13} x_{21} x_{22} x_{23} x_{31} x_{32} x_{33} t_1 t_2 t_3 :** ermöglicht das explizite Setzen einer affinen Transformation des resultierenden Datensatzes; x_{ij} beschreibt die Rotationsmatrix und t_i den Translationsvektor,
- alignWith x_{11} x_{12} x_{13} x_{21} x_{22} x_{23} x_{31} x_{32} x_{33} t_1 t_2 t_3 :** ermöglicht das explizite Setzen einer affinen Transformation für `in2.txt`; x_{ij} beschreibt die Rotationsmatrix und t_i den Translationsvektor,
- addUnmatched:** fügt Trajektorien von `in2.txt` auch dann dem resultierenden Datensatz `out.txt` hinzu, wenn für diese keine korrespondierenden Trajektorien in `in1.txt` gefunden wurden,
- delay frames:** statt der standardmäßigen automatischen Suche des zeitlichen Versatzes bei der Eingabedatensätze, gibt `frames` explizit an, um wieviele Einzelbilder `in1.txt` nach `in2.txt` beginnt,
- varyingHeight:** erlaubt eine variierende Personengröße bzw. z -Wert in der Ausgabedatei.



Kommerzielle Produkte zur Personenerfassung

Aufgrund von vielfältigen kommerziellen Anwendungen (siehe u. a. Abschnitt 2.1) gibt es eine Reihe von Firmen, die Produkte zur Erfassung von Personen anbieten. Folgende Listen wurden bei der Suche nach Produkten für Feldstudien und bei Messebesuchen zusammengetragen und enthalten neben dem Firmennamen einen Produktnamen, wenn die Firma nicht ausschliesslich Systeme zur Personenerfassung anbietet. Die meisten Systeme arbeiten mit Bildern von sichtbarem Licht erfassenden monokularen Kameras, da diese kostengünstig sind und die Nutzung von bereits installierten Überwachungskameras ermöglichen. Ist bekannt, dass auch andere Techniken (u. a. Laserscanner, Stereokameras, Tiefenkameras, Wärmebildkameras) eingesetzt werden, werden diese Firmen in speziellen Listen geführt.

- Visiosafe
<http://www.visiosafe.com/>
- Cognimatics, TrueView People Counter
<http://www.cognimatics.com/products/people-counter>
- Data Components
<http://www.kundenzaehlen.de/prodTVPC.html>
- Video Turnstile
<http://videoturnstile.com/>
- Acorel, automatic people counting solution
<http://www.acorel.com/>
- VCA Technology
<http://www.vcatechnology.com/>

- AIT, smart systems mit smart eye
<http://www.ait.ac.at/research-services/research-services-safety-security/new-sensor-technologies/entwicklung-von-embedded-systemen-fuer-kundenspezifische-loesungen/smart-eye-ucos-universal-counting-sensor>
- ISL, European Laboratory for Sensory Intelligence (ELSI), Pedestrian Flow Analyser
<http://www.isl.eu/index.php/de/elsi/pedestrian-flow-analyser-new>
- Vaelsys, ViFende3, Modul ViconC
<http://vaelsys.com/our-products>
- Visual Tools, AX-PeCo
<http://www.visual-tools.com/en/products/axis-compatible-products-2/people-counting-application-ax-peco>
- intuVision, u.a. Line Crossing Spy
http://intuvisiontech.com/products/edge_analytics_LineCrossing.php
- Acic, ActivityDetection
<http://www.acic.eu/en/products/mv-activity-detection.html>
- 3Y Teknoloji, People counter
<http://www.3yteknoji.com.tr/en/applications/retail/people-counter.html>
- TechnoAware, VTrack
<http://www.technoaware.com/eng/products>
- AllGoVision, AllGoVision Edge Analytics
http://www.allgovision.com/agv_edge_analytics.php
- Agent Vi, Vi-System
http://www.agentvi.com/61-Products-62-Vi_System
- A.I. Tech, AI-PEOPLE
<http://www.aitech-solutions.eu/en/ai-people>
- Digital Barriers, SafeZone-edge
<http://www.digitalbarriers.com/safezone-edge>
- Foxstream, Foxcounter
<http://www.foxstream.fr/newsite/en/products.html#produits-counter>
- Mindtree, Video Analytics Software (VAS)
<http://www.mindtree.com/solutions/digital-video-surveillance/video-analytics-software>

- **IPS Intelligent Video Analytics, IPS Motion Detection**
<http://www.ips-analytics.com/en/products/ips-videoanalytics-new/camera-based/ips-motion-detection-axis.html>
- **AXIS Communication, AXIS Digital Autotracking**
<http://www.axis.com/de/products/autotracking/index.htm>

Stereokamera:

- **Visapix**
<http://www.visapix.de/>
- **Vitracom (neben normalen, auch Zeilen- und Tiefenkameras)**
<http://www.vitracom.de/>
- **PointGrey, BumbleBee-Kameras mit Censys3D**
<http://www.ptgrey.com/products/censys3D>
- **Infodev (neben normalen, auch Zeilen- und Tiefenkameras)**
<http://www.infodev.ca/>

Laserscanner:

- **nvl², MoPeCo**
<http://www.nvlhoch2.de/>
- **LASE PeCo Systemtechnik GmbH**
<http://www.peoplecounter.de/>
- **LASE Industrielle Lasertechnik GmbH, LTS 400**
<http://www.lase.de/produkte/gebaeudeautomation/laser-tracking-system.html>

Wärmebildkamera:

- **Bosch, Intelligente Video Analysis (IVA) mit BEV-Personenzähler, Wärmebildkamera in MIC Serie 612**
http://de.boschsecurity.com/de/master_product/02_products_1/st_bu_f_277305_nsode_catalog_prod_de/st_section_f_277673_nsode_catalog_prod_de/st_prodfam_p_277673_nsode_catalog_prod_de_289986
- **Irisys, PeopleCounter**
<http://www.irisys.co.uk/people-counting>
- **TRAF-SYS**
<http://www.trafsys.com/people-counting>

Tiefenkamera:

- Sick, TVS100 Vision-Kamera
http://www.sick.com/group/DE/home/products/product_news/laser_measurement_systems/Seiten/tvs100_vision_camera.aspx
- Abtek Controls
<http://www.abtekcontrols.com/Principals/People/People.htm>
- Daviko (noch in der Forschung, Nutzung des Kinect-Sensors)
<http://www.daviko.com/>
<http://www.cebit.de/produkt/crowd-detector/471080/J937606>



Nomenklatur

Englische Übersetzungen werden bei wichtigen Begriffen kursiv in Klammern dahinter geschrieben, da die Namen der Bezeichner davon abgeleitet werden oder die Begriffe in der überwiegend englischsprachigen Literatur verbreitet sind.

Mathematische Größen werden kursiv und wie folgt geschrieben:

skalar klein
vektor klein mit Pfeil
MATRIX groß

Vektoren werden als Spaltenvektoren definiert und die Indizierung läuft wie folgt:

$$\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = (a_1 \ a_2 \ a_3)^T, \quad A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Natürliche Zahlen beginnen bei Eins ($\mathbb{N} = \{1, 2, 3, \dots\}$) und eine Teilmenge der ersten N Elemente wird mit $\mathbb{N}_N = \{1, 2, 3, \dots, N\}$ bezeichnet.

Die Standardabweichung von N Messgrößen $a_i, i \in \mathbb{N}_N$ ist über die korrigierte Stichprobenvarianz definiert als

$$\sqrt{\frac{1}{N-1} \sum_{i=1}^N \left(a_i - \frac{1}{N} \sum_{i=1}^N a_i \right)^2}$$

und wird mit \pm angegeben.

F.1 Symbolverzeichnis

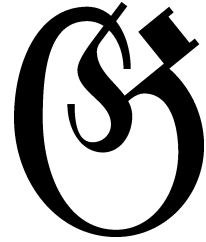
Liste der in dieser Arbeit verwendeten Symbole, die wiederholt eingesetzt werden. Die Seitenzahl der zweiten Spalte gibt an, auf welcher Seite das Symbol eingeführt oder das erste Mal benutzt wurde:

\vec{a}		Vektor \vec{a} erweitert um homogene Koordinate: $(a_1 \ a_2 \ a_3 \ w)^T$
$\bar{a} = \varnothing a$		arithmetisches Mittel von a_i (z. B. Durchschnittspersonengröße \bar{p}^h)
$\vec{u} = (u \ v)^T$	48	Bildpunkt bzw. Pixel mit horizontaler und vertikaler Koordinate u, v [px]
$\vec{x} = (x \ y \ z)^T$	48	Position im realen Raum [m]
$\vec{x}' = (x' \ y' \ z')^T$	52	Position im Kamerakoordinatensystem [m]
\vec{t}	51	Translationsvektor (Lage einer Kamera) [m]
R	51	Rotationsmatrix (Blickrichtung einer Kamera)
f bzw. f_x, f_y	51	Brennweite (<i>focal length</i>) einer Kamera [m] (f' Brennweite in Pixel [px])
s bzw. s_x, s_y	51	Skalierungsfaktor (Umrechnungsfaktor Pixelwerten in metr. Maße) [m/px]
\vec{v}	51	Fluchtpunkt (<i>vanishing point</i>) [m]
$\vec{c} = (c_x \ c_y)^T$	51	Kamerahauptpunkt (<i>principal point</i>) [px] in der Bildebene
k_1, k_2	51	radiale Verzerrungsparameter
p_1, p_2	51	tangentiale Verzerrungsparameter
$p'_i(t)$	69	Ort in der Bildebene [px] von Person i zum Zeitpunkt t
p'_i	69	Laufweg in der Bildebene von Person i
p'	69	Menge aller Laufwege in der Bildebene
$p_i(t)$	82	Ort im realen Raum [m] von Person i zum Zeitpunkt t
p_i	82	Laufweg im realen Raum von Person i
p	82	Menge aller Laufwege im realen Raum
p_i^h	83	Personengröße [m] von Person i
\bar{p}^h	83	Durchschnittspersonengröße [m]
a	83	Aufhängehöhe [m] einer Kamera (Abstand zum Boden)
N	69	Anzahl von Personen
t_i^{in}, t_i^{out}	69	Ein- und Austrittszeit [s] der Person i in einer Bildsequenz
$B_i(u, v)$	69	Bildpunkt eines Bildes B_i an Pixelkoordinate $(u \ v)^T$ einer Bildsequenz B
B^L	70	Level L einer Gauß-Pyramide von Bildsequenz B
F	69	Frameanzahl einer Bildsequenz
d_p	187	Personenabstand [m]
d_h	187	Größendifferenz [m] von Personen
w_k	187	Kopfausdehnung [m]
w_m	187	Ausdehnung [m] des Bereichs, der zur Erk. eines Markers einsehbar sein muss
w_s	103	Schulterbreite [m] einer Person
p_h	192	horizontale Sensorauflösung [px]
p_m	192	zur Erkennung benötigte Ausdehnung [px] des Markers auf dem Sensor
b	88	Basislänge [m] auf Basislinie (<i>baseline</i>) eines Stereokamerasystems
$\delta(u, v)$	95	Disparität (<i>disparity</i>) [px] des Bildpunktes $(u \ v)^T$
C	136	Optisches Zentrum [m]
T'	102	Matrix aller Disparitäten $\delta(u, v)$ eines Bildes
T	102	Perspektivisches metrisches Tiefenfeld basierend auf T'
H	102	Perspektivisches metrisches Höhenfeld basierend auf T und Aufhängehöhe a
H_{bg}	102	perspektivisches Höhenfeld des Hintergrundes

r_1, r_2	112	Längen [m] der kleinen und großen Halbachse einer Ellipse
$a_i(t_j)$	162	Betrag der mikroskopischen Beschleunigung [m/s ²] von Person i zum Zeitpunkt t_j
e_i	69	Farbdifferenz von Bildbereichen während des Trackings
e'_i	75	Flächenbereinigte Farbdifferenz e_i während des Trackings
q'_i	75	Qualität eines Tracking-Schrittes basierend auf e'_i
e_h	83	Positionsfehler [m] aufgrund unterschiedlicher Personengrößen und persp. Sicht
e_c	136	Positionsfehler [m] des Farbmarkers aufgrund der persp. Sicht
$e_{\text{Verfahren}}$	141	Positionsfehler [m] der markerlosen Personendetektion aufgrund von „Verfahren“
e_o	177	Positionsfehler [m] des Verfahrens ohne Marker aufgrund der persp. Sicht

F.2 (zusammengesetzte) Einheiten

[px]	Pixel
[bit]	Bit
[f/s]	Bilder Pro Sekunde (<i>frames per second</i>) bzw. Bildwiederholrate, oft auch [fps]
[bpp]	Bits pro Pixel (<i>bits per pixel</i>) oder auch [bit/px]
[par]	Pixelseitenverhältnis (<i>pixel aspect ratio</i>)



Glossar

Das Glossar führt wichtige in dieser Arbeit auftretende Begriffe thematisch gruppiert in der Reihenfolge ihrer Erscheinung auf und soll insbesondere das fächerübergreifende Verständnis der Arbeit erhöhen. Nicht eindeutig definierte Begriffe werden gemäß Verwendung in dieser Arbeit angegeben.

Kamera: bezeichnet ohne weitere Angaben eine monokulare Apparatur, die vom menschlichen Auge wahrnehmbare elektromagnetische Wellen mittels eines 2D-Sensors über eine Optik als perspektivisches Abbild der Umgebung erfasst,

Kamerakalibrierung: Berechnung der Abbildungseigenschaften der Kamera wie der extrinsischen und intrinsischen Parameter u. a. zur Bildmessung bzw. Photogrammetrie,

extrinsische Parameter: beschreiben die Abbildung des Welt- in das Kamerakoordinatensystem und legen damit die **äußere Orientierung** der Kamera (**Rotation, Translation**) fest,

intrinsische Parameter: beschreiben die Abbildung des Kamera- in das Bildkoordinatensystem und legen damit die **innere Orientierung** der Kamera fest; die Parameter enthalten den Abstand der **Bildebene** zum Ursprung des Kamerakoordinatensystems (**Brennweite**), den Umrechnungsfaktor zwischen Pixelwerten in metrische Maße, den Schnittpunkt der optischen Achse mit der Bildebene (**Kamerahauptpunkt**) und je nach Kameramodell unterschiedliche **Verzerrungsparameter** (z. B. radiale, tangentielle),

optische Achse: ist die Symmetrieachse des Linsensystems der Optik einer Kamera und verläuft durch den Kamerahauptpunkt in der Bildebene; im optimalen Fall steht die optische Achse senkrecht und mittig auf dem Bildsensor und der Kamerahauptpunkt liegt in der Bildmitte,

Bildwinkel: gibt den maximal einzusehenden Öffnungswinkel einer Kamera an (horizontal, vertikal oder diagonal) und ist abhängig von der Brennweite und Größe des Bildsensors,

Bildkoordinatensystem: zweidimensionales kartesisches Koordinatensystem mit der sensorspezifischen Einheit Pixel; die Koordinatenachsen verlaufen parallel zu den Bild- bzw. Sensorzeilen und -spalten; der Ursprung liegt entweder in der unteren linken Bild- bzw. Sensorecke und das

Koordinatensystem spannt ein rechtshändiges oder in der oberen linken Ecke und spannt ein linkshändiges Koordinatensystem auf,

Kamerakoordinatensystem: dreidimensionales rechtshändiges kartesisches Koordinatensystem, dessen z-Achse der optischen Achse entspricht und die beiden anderen Achsen parallel zu den Achsen des Bildkoordinatensystems liegen; der Ursprung befindet sich im **optischen Zentrum** des Linsensystems der Kamera, die Größe der Brennweite entfernt von der Bildebene; die Einheit entspricht der des Weltkoordinatensystems,

Weltkoordinatensystem: dreidimensionales rechtshändiges kartesisches Koordinatensystem des realen Raumes zu dem alle Objekte in ihrem lokalen Koordinatensystem relativ positioniert werden, so z. B. die Kamera über ihre extrinsischen Parameter,

geometrische Verzerrung: bzw. **Verzeichnung** eines optischen Systems ist die für eine Vermessung wichtigste Aberration; der Abbildungsfehler entsteht durch das Linsensystem einer Kamera und beinhaltet u. a. **radiale und tangentiale Verzerrungen** wie z. B. tonnen- oder kissenförmige Verzeichnungen, die über intrinsische Parameter eines **Kameramodells** angegeben und bei der Transformation vom Bild- in das Kamerakoordinatensystem berücksichtigt werden,

perspektivische Verzerrung: bezeichnet die durch die **perspektivische Transformation** des dreidimensionalen Raumes auf die Bildebene entfernungsabhängige Größenänderung der abgebildeten Objekte,

Isolinie: Linie, entlang deren der gleiche Wert (Isowert) vorliegt; **gerichtete Isolinie** bezeichnet eine Isolinie, deren Orientierung bzw. Umlaufrichtung die Lage der dem Isowert kleiner und größeren Werte angibt,

Ellipse: Schrägbild eines Kreises, das die Punkte einer Ebene beschreibt, deren Summe der Abstände zu zwei Punkten (**Brennpunkte**) identisch ist; der Mittelpunkt liegt in der Mitte der Brennpunkte; die innenliegende Achse durch die Brennpunkte heißt **Hauptachse**; der Mittelpunkt teilt die Hauptachse in die **großen Halbachsen** der Länge r_1 ; die senkrecht zur Hauptachse durch den Mittelpunkt verlaufende Achse heißt **Nebenachse** und wird in zwei **kleine Halbachsen** der Länge r_2 geteilt; die **numerische Exzentrizität** $\sqrt{r_1^2 - r_2^2} / r_1 \in [0, 1)$ sagt etwas über das Seitenverhältnis der Haupt- und Nebenachse aus,

HSV-Farbraum: dreidimensionaler Farbraum, der die Farbe über den **Farbton**, die **Sättigung**, und die **Helligkeit** spezifiziert und sich zur natürlichen Farbwahl von einem Menschen eignet,

Öffnen, Schließen: im Bereich der Bildverarbeitung morphologische Basis-Operationen u. a. zur Beseitigung von Störungen und zur Glättung von Bildstrukturen durch eine aufeinanderfolgende Ausführung einer **Dilatation** (Ausdehnung einer Struktur) und **Erosion** (Beschneidung einer Struktur),

optischer Fluss: gibt die Bewegung von Bildpunkten in der Bildebene zwischen zwei Bildern einer Bildsequenz als Vektorfeld an,

Hintergrundsubtraktion: Segmentierung des statischen Hintergrundes eines Bildes, um für nachgelagerte Verarbeitungsschritte nur den Vordergrund betrachten zu können,

Stereokamera: besitzt in der Regel zwei synchron arbeitende nebeneinander angeordnete Kameras zur Erfassung der gleichen Szene, um entsprechend dem räumlichen Sehen des Menschen über den Abstand der Pixel der Abbildungen beider Kameraperspektiven die Entfernung des im Pixel abgebildeten Punktes über eine Triangulation zu bestimmen; die **Basislänge** ist der Abstand der optischen Zentren der zwei Kameras,

Epipolargeometrie: beschreibt die geometrische Beziehung zwischen den beiden Ansichten einer Stereokamera durch eine **Fundamentalmatrix** und erlaubt dadurch die Ausrichtung der Kameraperspektiven mittels **Rektifizierung**, so dass gleiche Objekte auf gleiche Bildzeilen abgebildet werden,

Triangulation: geometrische Methode zur optischen Abstandsmessung durch Winkelbestimmung im Dreieck, das von zwei bekannten Betrachtungspunkten und dem Objektpunkt, zu dem die Entfernung bestimmt werden soll, aufgespannt wird; im Gegensatz dazu bestimmen **Laufzeitverfahren** die Entfernung über die Dauer, die ein Signal für den Weg zum Objekt und zurück braucht,

Disparität: Abstand der Pixelpositionen der den gleichen Objektpunkt abbildenden Pixel innerhalb gleicher Bildzeilen der Abbildungen beider Kameraperspektiven einer Stereokamera; die **Disparitätsmatrix** enthält alle bestimmbaren Disparitätswerte eines Bildes einer Stereokamera; das **perspektivische Tiefenfeld** gibt je Pixel die Entfernung des Objektes zur Kamera entlang der optischen Achse auf dem Sehstrahl durch das jeweilige Pixel und dem optischen Zentrum an; die Tiefe ist umgekehrt proportional zur Disparität,

Tiefenauflösung: Genauigkeit der Tiefeninformation; sie ist für die Abstandsmessung mittels Triangulation abhängig von der Entfernung zum Sensor,

Tiefenbereich: Intervall zwischen minimal und maximal bestimmbarer Tiefe einer Technik zur Abstandsmessung,

Feldstudie: Beobachtungen im natürlichen Umfeld ohne jegliche Beeinflussung der Umgebung und der Studienobjekte bzw. Personen; im Gegensatz zum **Feldexperiment**, bei dem im natürlichen Umfeld interessierende Parameter verändert werden, und dem **Laborexperiment**, bei dem das gesamte Umfeld künstlich hergestellt wird und die Studienobjekte sich der Teilnahme bewusst sind,

Fehlerrate: zur Beurteilung eines **Klassifikators**, z. B. einer Methode zur Detektion einer Person; die **Richtig-Positiv-Rate** gibt den Anteil der korrekt als positiv klassifizierten Objekte an der Gesamtheit der tatsächlich positiven Objekte an; die **Falsch-Negativ-Rate** ist der Anteil der **falsch negativen**, also fälschlich als negativ klassifizierten Objekte, die in Wirklichkeit positiv sind; die **Richtig-Negativ-Rate** gibt den Anteil der korrekt als negativ klassifizierten Objekte an der Gesamtheit der in Wirklichkeit negativen Objekte an; die **Falsch-Positiv-Rate** ist der Anteil der **falsch positiven**, also fälschlich als positiv klassifizierten Objekte, die in Wirklichkeit negativ sind, an der Gesamtheit der in Wirklichkeit negativen Objekte; der **positive Vorhersagewert** gibt den Anteil der korrekt als positiv erkannten Objekte an der Gesamtheit der als positiv erkannten Objekte an; die **Falsch-Ermittlungs-Rate** ist der Anteil der **falsch positiven** an der Gesamtheit der tatsächlich als positiv klassifizierten Objekte,

- makroskopisches Modell:** beschreibt die zeitliche und räumliche Bewegung eines Personenstroms, indem dieser als eine Einheit mit konstanten Maßen wie z. B. einer globalen Dichte aufgefasst wird,
- mikroskopisches Modell:** beschreibt die Bewegung jedes Einzelnen mit individuellen Eigenschaften wie z. B. einer Wunschgeschwindigkeit; hierdurch sind komplexe Bewegungen von Menschenmassen, wie z. B. kreuzende oder gegenläufige Ströme, abbildbar; typische Modelle sind **Zellularautomaten-Modelle** und Modelle basierend auf **sozialen Kräften**,
- Trajektorie:** oder auch **Laufweg** ist in dieser Arbeit der Weg über die Zeit t , den der Kopf einer Person i zurücklegt; dieser kann in der Bildebene in Pixelkoordinaten ($p'_i(t)$) oder im realen Raum ($p_i(t)$) angegeben werde,
- Geschwindigkeit:** der zurückgelegte Weg pro Zeit Δt als Betrag lautet $v_i(t) = \|p_i(t + \Delta t) - p_i(t - \Delta t)\| / (2\Delta t)$; geglättete Laufwege ermöglichen die reine Berücksichtigung der Hauptbewegungsrichtung,
- Beschleunigung:** der Betrag der **mikroskopische Beschleunigung** $a_i(t_j) = \|(p_i(t_{j+1}) - p_i(t_j)) - (p_i(t_j) - p_i(t_{j-1}))\| / (\Delta t)^2$ wird als Maß für die Ruhe innerhalb einer Trajektorie herangezogen,
- Dichte:** Anzahl der Personen N innerhalb einer Menschenmenge pro Flächeneinheit A : $\rho = N/A$; die **Voronoi-Dichte** erlaubt eine gering fluktuierende, räumlich detaillierte Berechnung der Dichte auf Basis von den persönlichen Platz beschreibenden Voronoi-Zellen,
- Fluss:** gibt den Durchsatz an Personen N pro Zeitintervall Δt an einer Kreuzungslinie an: $J = N/\Delta t$; der **spezifische Fluss** normiert den Fluss im Bezug auf die Länge w der Kreuzungslinie: $J_s = J/w$; die **Kapazität** einer Fußgängeranlage ist der maximale Fluss durch die Anlage (Maximum im Dichte-Fluss-Diagramm),
- Fundamentaldiagramm:** beschreibt als wichtigstes quantitatives Kennzeichen einer Fußgängeranlage im Verkehrswesen und damit in der Fußgängerdynamik den Zusammenhang zwischen Dichte und Fluss im **stationären Zustand** (sich im Gesamtsystem kaum verändernde Geschwindigkeit und Dichte) als **Dichte-Fluss-Diagramm** $J_s(\rho)$; aufgrund der Abhängigkeit $J_s = \rho \cdot v$ kann das Diagramm auch als **Dichte-Geschwindigkeits-Diagramm** $v(\rho)$ formuliert werden,
- Voronoi-Diagramm:** unterteilt in der Fußgängerdynamik die Bewegungsebene beschränkt auf die Fußgängeranlage eindeutig in so genannte **Voronoi-Zellen**, die die Ebene in Bereiche zerlegen, die jeder Person das ihr gemäß euklidischem Abstand am nächsten liegende Flächenstück zuweist, das damit als persönlicher Platzbedarf aufgefasst werden kann,
- Geschwindigkeits-, Dichte-, Fluss-Profil:** für die gesamte Fußgängeranlage über die Zeit eines stationären Zustandes integrierte Geschwindigkeit, Dichte oder integrierter Fluss; da deshalb für die gesamte Fläche jederzeit ein Ausgangswert definiert sein muss, bietet sich die Anwendung der Voronoi-Maße der Geschwindigkeit, Dichte und des Flusses an.

Band / Volume 15

UNICORE Summit 2012

Proceedings, 30 - 31 May 2012 | Dresden, Germany

edited by V. Huber, R. Müller-Pfefferkorn, M. Romberg (2012), iv, 143 pages

ISBN: 978-3-89336-829-7

URN: urn:nbn:de:0001-2012111202

Band / Volume 16

**Design and Applications of an Interoperability Reference Model
for Production e-Science Infrastructures**

by M. Riedel (2013), x, 270 pages

ISBN: 978-3-89336-861-7

URN: urn:nbn:de:0001-2013031903

Band / Volume 17

**Route Choice Modelling and Runtime Optimisation
for Simulation of Building Evacuation**

by A. U. Kemloh Wagoum (2013), xviii, 122 pages

ISBN: 978-3-89336-865-5

URN: urn:nbn:de:0001-2013032608

Band / Volume 18

Dynamik von Personenströmen in Sportstadien

by S. Burghardt (2013), xi, 115 pages

ISBN: 978-3-89336-879-2

URN: urn:nbn:de:0001-2013060504

Band / Volume 19

Multiscale Modelling Methods for Applications in Materials Science

by I. Kondov, G. Sutmann (2013), 326 pages

ISBN: 978-3-89336-899-0

URN: urn:nbn:de:0001-2013090204

Band / Volume 20

**High-resolution Simulations of Strongly Coupled Coulomb Systems
with a Parallel Tree Code**

by M. Winkel (2013), xvii, 196 pages

ISBN: 978-3-89336-901-0

URN: urn:nbn:de:0001-2013091802

Band / Volume 21

UNICORE Summit 2013

Proceedings, 18th June 2013 | Leipzig, Germany

edited by V. Huber, R. Müller-Pfefferkorn, M. Romberg (2013), iii, 94 pages

ISBN: 978-3-89336-910-2

URN: urn:nbn:de:0001-2013102109

Band / Volume 22

**Three-dimensional Solute Transport Modeling in
Coupled Soil and Plant Root Systems**

by N. Schröder (2013), xii, 126 pages

ISBN: 978-3-89336-923-2

URN: urn:nbn:de:0001-2013112209

Band / Volume 23

**Characterizing Load and Communication Imbalance
in Parallel Applications**

by D. Böhme (2014), xv, 111 pages

ISBN: 978-3-89336-940-9

URN: urn:nbn:de:0001-2014012708

Band / Volume 24

**Automated Optimization Methods for Scientific Workflows in e-Science
Infrastructures**

by S. Holl (2014), xvi, 182 pages

ISBN: 978-3-89336-949-2

URN: urn:nbn:de:0001-2014022000

Band / Volume 25

**Numerical simulation of gas-induced orbital decay of binary systems
in young clusters**

by A. C. Korntreff (2014), 98 pages

ISBN: 978-3-89336-979-9

URN: urn:nbn:de:0001-2014072202

Band / Volume 26

UNICORE Summit 2014

Proceedings, 24th June 2014 | Leipzig, Germany

edited by V. Huber, R. Müller-Pfefferkorn, M. Romberg (2014), iii, 60 pages

ISBN: 978-3-95806-004-3

URN: urn:nbn:de:0001-2014111408

Band / Volume 27

**Automatische Erfassung präziser Trajektorien
in Personenströmen hoher Dichte**

by M. Boltes (2015), xii, 308 pages

ISBN: 978-3-95806-025-8

URN: urn:nbn:de:0001-2015011609

Simulationen können helfen, Verkehrsanlagen für Fußgänger komfortabel und sicher zu gestalten. Das Verständnis über die Fußgängerdynamik ist dabei wesentlich für die Entwicklung verlässlicher Modelle. Hierfür sind detaillierte und reproduzierbare Daten realer Bewegungen von Menschenmassen und Individuen nötig, um das Bewegungsverhalten zu analysieren, daraufhin Modellideen zu entwickeln, deren Umsetzung zu kalibrieren und am Ende das Ergebnis zu validieren.

Diese Arbeit beschreibt die Erhebung solcher Bewegungsdaten, die aus Überkopfaufnahmen von Kameras gewonnen werden. Dabei werden individuelle Laufwege jeder Person extrahiert. Mit Hilfe dieser Trajektorien werden die wichtigsten Messgrößen zur Beschreibung der Dynamik von Menschenströmen wie u. a. die Personendichte, die Geschwindigkeit und damit der Personenfluss bestimmt.

Besonders bei Stau und Gedränge ist mit kritischen Situationen zu rechnen. Daher ist es für die zur Extraktion der Laufwege entwickelten Verfahren wichtig, dass diese auch für hohe Personendichten verlässlich anwendbar sind.

Die Bewegung von Fußgängern wird von vielen Faktoren wie z. B. der Geometrie, der Personendichte, der Motivation und dem Kulturkreis beeinflusst. Zur Untersuchung dieser Vielzahl an Einflussfaktoren sind eine große Anzahl an Versuchen mit zum Teil hohen Personenzahlen durchzuführen. Die automatische Erfassung der Laufwege bietet im Vergleich zur manuellen Erfassung deutliche Vorteile hinsichtlich des Zeitaufwandes, der Genauigkeit und der Reproduzierbarkeit.

Die Prozesskette der Extraktion besteht aus der Kalibrierung der Bilder, der Detektion der Personen, dem Tracking und der Bestimmung des realen Aufenthaltsortes. Für die Detektion einer Person wurden verschiedene den unterschiedlichen Anwendungsfällen und örtlichen Gegebenheiten angepasste Marker mit entsprechenden Erkennungsverfahren entwickelt. Insbesondere für Feldstudien ist zudem ein markerloses Verfahren erarbeitet worden, das mit Hilfe von Stereokameras auch in dichten Menschenmassen eine hohe Erkennungsrate aufweist. Alle Entwicklungen sind in die den gesamten Prozess abbildende Software PeTrack eingeflossen. Für die Nachvollziehbarkeit der Ergebnisse wird neben den Verfahren auch die technische Umsetzung der Erfassung beschrieben.

Diese Publikation wurde vom Jülich Supercomputing Centre (JSC) herausgegeben, einem Teilinstitut des Institute for Advanced Simulation (IAS). Das IAS vereint Simulationswissenschaften und Supercomputing unter einem Dach. Es umfasst diejenigen wissenschaftlichen Institute des Forschungszentrums Jülich, die Simulationen auf Supercomputern als vorrangigste Forschungsmethode verwenden.